HELSINKI UNIVERSITY OF TECHNOLOGY
Faculty of Electronics, Communications and Automation

Markus Wallasvaara

An Evaluation Assessment of a Vowel Training System: The Vowel Game

Master's Thesis
Turku, 12th May 2008

Supervisor: Professor Paavo Alku
Instructor: Professor Olli Aaltonen

HELSINKI UNIVERSITY OF TECHNOLOGY
Faculty of Electronics, Communications and Automation

MASTER'S THESIS: AN ABSTRACT

Author: Markus Wallasvaara

Date: 12th May 2008

Pages: 98

Title: An Evaluation Assessment of a Vowel Training System: The Vowel Game

Department: Dept. of Signal Processing and Acoustics

Code: S-89

Supervisor: Professor Paavo Alku

Instructor: Professor Olli Aaltonen

The purpose of this thesis was to study the effects of real-time visual feedback on the pronunciation/imitation of the Finland Swedish vowel /ʉ/ and to answer the research question of whether or not visual feedback has an effect in making the pronunciation more constant. The center of the study was the new vowel training system, The Vowel Game, which was implemented by Annu Paganus at the department of Information Technology at the University of Turku. The Vowel Game was tested using 20 Finnish adults aged 19-30 yrs. who had no Swedish-speaking background other than that studied at school. The test subjects were divided into two groups. The treatment group, which consisted of 6 male and 4 female subjects, used The Vowel Game's visual real-time feedback under controlled circumstances in imitating the sustained /ʉ/-vowel and a word containing this vowel /hʉs/ ('hus' meaning 'house') according to both synthetic (/ʉ/) and natural (/hʉs/) audio models. The control group, also consisting of 6 male and 4 female subjects, was never exposed to the effects of The Vowel Game's visual feedback, but only imitated the /ʉ/-vowel and the 'hus'-word under the same circumstances using the same audio models as the treatment group. Their imitations were nevertheless recorded with The Vowel Game. The hypothesis was that the treatment group's deviation in the pronunciation of the vowels would be smaller in the end in comparison to the control group's performance. The analysis of the results strongly indicated that The Vowel Game's real-time visual feedback does have a positive effect on the constancy of the pronunciation of the /ʉ/-vowel. This would suggest that The Vowel Game has the potential of becoming a tool in speech therapy and a help to individuals attempting to learn new vowel categories. The study also takes a look at some of the previous vowel training systems and their main differences with The Vowel Game alongside the mathematical methods of formant retrieval and the features and characteristics of the vocal tract. In addition, the study presents the general views on the perception and categorization of speech sounds as the theoretical background for The Vowel Game.

Keywords:
Real-time visual feedback, imitation, pronunciation, vowels, vowel diagram, formants, linear predictive coding.

TEKNILLINEN KORKEAKOULU
Elektroniikan, tietoliikenteen ja automaation tiedekunta

DIPLOMITYÖ:
TIIVISTELMÄ

Tekijä: Markus Wallasvaara	Päivämäärä: 12.5.2008 Sivumäärä: 98
Diplomityön nimi: An Evaluation Assessment of a Vowel Training System: The Vowel Game	
Laitos: Signaalinkäsittelyn ja akustiikan laitos	Koodi: S-89
Valvoja: Professori Paavo Alku	
Ohjaaja: Professori Olli Aaltonen	

Tämän diplomityön tarkoituksena oli selvittää, onko reaaliaikaisella visuaalisella palautteella vaikutusta suomen ruotsin /ʉ/-vokaalin matkimiseen, ja voiko tämän kaltainen palaute tehdä lausumisesta säännöllisempää. Tutkimuksen kohteena oli Turun Yliopiston Informaatioteknologian laitoksella luotu, Annu Paganuksen toteuttama The Vowel Game-niminen vokaalipeli. Vokaalipeliä testattiin 20 suomalaisella aikuisella, joiden iät vaihtelivat välillä 19-30 vuotta ja joilla ei ollut entuudestaan ruotsinkielistä taustaa koulussa opitun ruotsin lisäksi. Henkilöt jaettiin kahteen ryhmään siten, että molemmissa oli 4 naista ja 6 miestä. Varsinainen testiryhmä hyödynsi Vokaalipelin reaaliaikaista visuaalista palautetta kontrolloiduissa testiolosuhteissa matkiakseen yksittäistä suomen ruotsin /ʉ/-vokaalia sekä sanaa /hʉs/ ’hus’ (talo). Matkiminen tapahtui tallenteelta kuultujen sekä synteettisten (/ʉ/) että luonnollisten (/hʉs/) mallien perusteella. Verrokkiryhmä matki ainoastaan malleja ilman pelin visuaalista palautetta. Tutkimusten hypoteesina oli olettamus, että visuaalista palautetta saanut testiryhmä kykenisi kohdistamaan imitoimansa vokaalit paremmin Vokaalipelin vokaalikartassa kuin verrokkiryhmä ja sitä kautta oppisi tuottamaan vokaalin säännöllisemmin. Tarkoituksena oli tutkia myös ryhmien keskimääräisten osumapisteiden hajontoja alku-ja lopputilanteissa, sekä tutkia, miten tulokset muuttuivat sekä ryhmien sisällä että ryhmien välillä. Tulosten analyysit osoittivat, että Vokaalipelin tuottama reaaliaikainen visuaalinen palaute auttoi testiryhmää selkeästi tuottamaan suomen ruotsin /ʉ/-vokaalin säännöllisemmin kuin verrokkiryhmä. Tämä tulos viittaisi siihen, että Vokaalipeliä voitaisiin käyttää työkaluna sekä puheterapiassa että apuvälineenä vieraan kielen opiskelijoille uusien vokaalikategorioiden oppimisessa. Tutkimuksessa esitetään katsaukset myös muutamiin aiempiin vokaalien lausumista opettaviin järjestelmiin sekä formanttianalyysiin ja sen matemaattisiin metodeihin. Lisäksi tutkimus esittää taustatietona perusteoriat puheentuottamisen anatomiasta sekä yleiset käsitykset puheäänten havaitsemisesta ja kategorisoimisesta.

Asiasanat:
Reaaliaikainen visuaalinen palaute, matkiminen, lausuminen, vokaalit, vokaalikartta, formantit, lineaariprediktio.

Preface

A work of this scope could not have been possible without the help of a multitude of individuals. Therefore, I wish to thank Professors Paavo Alku and Olli Aaltonen for their coaching and many insightful views and instructions on how to approach and compose this study. I also wish to thank Professor Tapio Salakoski and Pekka Reijonen at the department of Information Technology at the University of Turku for arranging all the necessary facilities at the ICT-building for my disposal. Without the crucial testing facilities, this work would have been nearly impossible. My utmost gratitude also goes to Annu Paganus, the creator of The Vowel Game. Without her creation and her contributions to the project, this work would never have existed in the first place. I also wish to thank Janne Savela, Stina Ojala, Lotta Alivuotila, Mikko Laakso and Ilkka Raimo for their many invaluable hints and tips in finally bringing this work to life. Of course, the many individuals who took part in the tests must not be forgotten. Without their contribution, I would still be only speculating. Finally, I wish to thank my mom and Jenni for being there when I needed it the most.

Turku, 12th May 2008

Markus Wallasvaara

Table of contents:

List of Figures:

List of Tables:

Abbreviations and symbols:

$\int$	The symbol for integration
α	The real part of the frequency response
β	The imaginary part of the frequency response
α_k	The linear predictor coefficient
$\frac{\partial}{\partial x}$	Partial derivative operator
A(z)	The transfer function of the LP inverse filter
Aveg.	Average
c	The speed of sound
c_n	The Fourier coefficient
C°	Celsius
CALL	Computer-assisted language learning
CAN	Communicative acquisitionist naturalistic
CAPT	Computer-assisted pronunciation training
CBST	Computer-based speech therapy
cont	Control
CoV	The Coefficient of variance
dB	Desibel
e(n)	The prediction error
E_n	The short-time average prediction error
ERB	Equivalent rectangular bandwidth
Eucl.	Euclidean
F	The formant of a neutral vowel
F_m	Formant value
f_s	The sampling frequency
F0	The fundamental frequency
F1	The first formant
F2	The second formant
F3	The third formant

F4	The fourth formant
F’2	A computation of F2, F3 and F4
G	Gain
Hz	Hertz
H(z)	The transfer function of the time-invariant digital filter
j	An imaginary unit
∞	Infinity
IPA	The International Phonetic Association
KTH	Kungliga Tekniska Högskolan
l	The length of the vocal tract
L	Lower matrix
lim	Limes
ln	Logarithm, e-based
log	Logarithm
LP	Linear prediction
LPC	Linear predictive coding
L1	The first language or mother tongue
L2	Second language
m	The values of the amplitude response
Mel	Mel-scale
ms	milliseconds
N	The amount of samples in the frequency response
ω	The angular velocity
p	p-value (statistical significance)
P-A-R	Pitch-Amplitude-Rhythm
P(z)	The system function of a linear predictor
SLA	Second language acquisition
StDev	Standard deviation
s(n)	Speech samples related to excitation
$s_n(m)$	A speech segment
S(z)	The output signal
t	time
tr	treatment

U	Upper matrix
u(n)	The Input signal in the time domain
U(z)	The Input signal in the frequency domain
/ʉ/	The Finland Swedish rounded upper back vowel
π	Pi
$\frac{V}{Hz}$	The Signal's spectral density
$\frac{\sqrt{W}}{Hz}$	The Signal's spectral density
w(m)	The Hann-window
x(t)	The Fourier series
X(f)	The Fourier transform
y(n)	The impulse response

1. Introduction

One of the most important key factors in oral communication is the clear and precise pronunciation of the different sounds of the spoken language. The probability that the listener does not identify the speaker's word or sentence correctly increases significantly when a phoneme or phonemes are mispronounced. At least such a mispronunciation will certainly draw the attention to the speaker's speech (Wiik 1998: 7). As an example, when a Chinese person attempts to pronounce the English word 'war', the result will most likely sound as /wo:l/ due to the fact that the Chinese sound system does not entail the consonant /r/. The sound is substituted with the consonant /l/ and therefore there is no distinction between how the English words 'wall' and 'war' would be pronounced according to the Chinese system of sounds. The likelihood that the uttered word is interpreted falsely as 'wall' instead of 'war' grows significantly in a situation where the context does not help the listener to identify the speaker's word correctly. Similarly, when a Japanese person attempts to pronounce the word 'English', the result will sound /Iŋriʃ/ ('Ingrish') due to the fact that the Japanese sound system does not include the consonant /l/ and the sound is substituted with the consonant /r/. The general idea in these examples is that pronunciation errors of the sounds of the target language will inevitably increase the difficulty of the listener to interpret the speaker's message correctly and thus hinder the concept of fluent communication. Also, pronunciation has a strong relationship with social power. If a student wishes to be accepted into the culture of the target language, proficient pronunciation and fluency are deemed crucial determiners of one's worth (Zhang 2004: 379).

One of the more traditional forms of training occurs when a student starts learning a new language at school. In most countries the emphasis in the learning process was usually placed on lexis and grammar. Relatively little attention was directed at how the student pronounced the different sounds of the target language (Zhang 2004: 378). Clear pronunciation errors were usually corrected, but within the confining time limits of the school system, 'the fine tuning' in pronunciation was often omitted altogether. Michael Carey argues that the role of pronunciation teaching in a second language (L2) has been controversial e.g. among English teachers. Some teachers believe it is crucial in communication whereas others tend to believe its importance is questionable (Carey 2002: 1).

While the above cases involving the Chinese and Japanese sound systems are but far-end examples of the difficulties in pronouncing English, the problems in learning pronunciation in a foreign language are common in all sound systems. This can be partially contributed to the problems related to the perception of speech sounds; a concept, which is examined more closely in the following chapters. Basically, the vowel sounds in a language are divided into vowel categories and when these categories of the student's mother tongue are different from the vowel categories of the target language, the perception and production of these new vowels may prove to be difficult. A student may often attempt to reproduce a foreign phoneme by using a phoneme of his/her own language that closely resembles the foreign phoneme. This obviously leads to a pronunciation error. Alternatively, the student may find it difficult to hear or reproduce a distinction between two phonemes in a case where the distinction does not exist in his/her mother tongue. All this appears to be typical of adult students. (Dowd, Smith and Wolfe 1998: 1). Down, Smith and Wolfe offer an example of the lack of distinction between /U/ and /y/ in English. The English, therefore, have difficulties in reproducing the distinction in French in words such as 'dessous' or 'dessus'. What adds to the problem is the fact that the number of vowels is often different in many languages. Still, there is always more than one vowel. As an example, classic Arabic has only three vowels (Pedersen, Rosenberg-Wolff and Uddström 1996b: 86).

It has been suggested that difficulties in pronunciation can be diminished or overcome with various forms of training alongside generic school education. For instance, CALL (Computer-aided language learning) and CAPT (Computer-assisted pronunciation training) are fields that have generated a considerable amount of interest and research in the past few decades. The Vowel Game is one of the most recent products under development to merge from these fields and it has been developed to improve the student's pronunciation of vowels utilising visual real-time feedback and a vowel diagram. The concept of feedback is considered very important in language learning, especially the concept of good-quality feedback (Zhang 2004: 380). In general, the idea of feedback in learning is deemed vital when the student's development and improvement are concerned.

In this study, The Vowel Game's effects and efficiency were evaluated by gathering results from Finnish students who had received education in the second domestic language in Finland; Swedish. The methods in the tests were based upon the concept of imitation. In a series of 6 consecutive tests and exercises, the test subjects were instructed to imitate the Finland Swedish vowel /ʉ/ and a word /hʉs/ ('hus' meaning 'house') containing that vowel, which were presented as pre-recorded audio models through a set of headphones. The reason why the vowel /ʉ/ was chosen as the center of the tests is due to the fact that it is the only sound in Finland Swedish that does not belong to the Finnish system of vowel sounds (Pedersen, Rosenberg-Wolff and Uddström 1996b: 88). It is therefore slightly more foreign to Finnish students. Taisto Määttä, for example, studied the perception of the vowel sounds of Swedish by speakers of Finnish. He concluded that the general problem that Finns have with regard to e.g. /ʉ/ is the general difficulty of learning rounded 'minus-back' vowels (1983: 199). The Finland Swedish /ʉ/ is indeed situated further towards the back of the perceptual vowel space of Finnish and is therefore slightly more difficult to learn.

In The Vowel Game's tests, the subjects were advised to imitate the audio models of the /ʉ/-vowel (alone and within a word) and to produce the vowel into The Vowel Game. The system consequently plotted the subject's vowel into a vowel diagram in real-time with respect to the first two formants of the sound. With this method, the subject received immediate feedback of where the imitation of the model sound was plotted in the diagram. After several imitations, the subjects began to see if the vowels were plotted in the same area or if they were scattered. The subjects were also given external guidance on how the pronunciation could be altered by changing the position and shape of the articulators (the lips and the tongue).

The feedback had to be easy to understand so that the subject learned how to improve the performance based on the feedback (Zhang 2004: 380). The idea behind the imitation of the audio model and the repetitions is based on the theory of B. F. Skinner who argued in 1957 that a stimulus will produce a reaction and that repetitions will reinforce the relationship between the stimulus and the reaction. Feedback will also contribute to the relationship (1957: 31). The concept of imitation is related also to the concept of the mirror-neuron system in humans, which according to Giacomo Rizzolatti and Laila Craighero, might explain the human capacity to learn by imitation (2004: 169). Michael Arbib also argues that the ability to imitate is a key in the evolutionary path leading to language with humans (2002: 272). Arbib, like Rizzolatti and Craighero, believes that the mirror neurons may explain speech imitation (Hurford 2004: 299). This concept is intriguing because the test subjects' performance is compared to a control group, which practices the pronunciation of the model sounds solely by imitation without the visual feedback of The Vowel Game. The comparison between the groups is performed to detect any short-time effects of the system and to determine if real-time visual feedback has an effect in making the pronunciation of the vowel more constant.

In the next chapter, the fundamental principles of speech processing, which form the basis for The Vowel Game, are given an overview along with the concepts of the vocal tract and how speech and sounds are perceived. Concepts such as formants, what they are and how they are retrieved in real-time from the speech signal are discussed in more detail. Some attention is also directed at the effects of age with regard to pronunciation. The signal processing methods such sampling, quantization and signal windowing are presented along with the principles of linear predictive coding (LPC). The autocorrelation method, which is the linear predictive analysis technique used in The Vowel Game, is given a brief overview alongside the fundamentals of the Fourier analysis whilst keeping in mind that the focus of this study is on the analysis of the student's results and on the evaluation of The Vowel Game's effects on learning pronunciation.

The history and the primary features of some of the previous vowel training systems are discussed in Chapter 4 along with how they differ from The Vowel Game. The actual details of the test methods concerning The Vowel Game are described in Chapter 5 and the results are presented in Chapter 6. Chapter 7 centers on the analysis of the results and finally Chapter 8 draws together the general conclusions of the test results and provides, as the title of this study states, an evaluation assessment of The Vowel Game.

2. Speech

This chapter gives a brief overview of the elements of the vocal tract, of how speech, consonants and vowels are formed and what articulatory phonetics is. The categorization of speech sounds is also examined since it is essentially linked with the testing of The Vowel Game.

2.1 The Vocal tract

Speech is unique to humans. No other living organism on the face of this earth is able to produce intelligent speech and therefore it is interesting to take a closer look at how speech is produced and how vowels are formed in the vocal tract. The speech processing mechanism is illustrated in Figure 1.

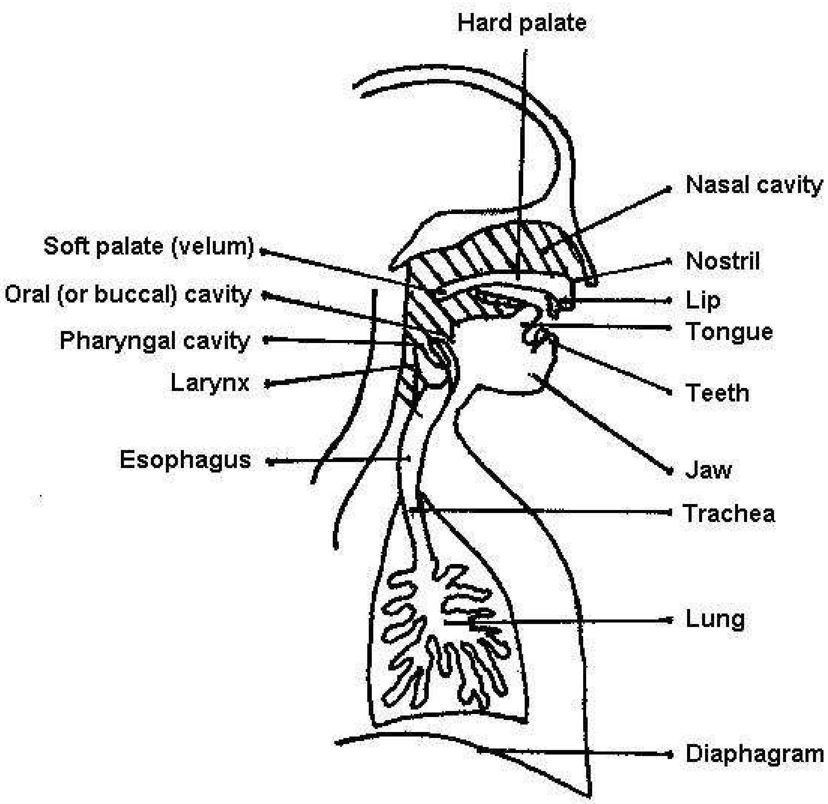


Figure 1 Speech production (modified from original) (Deller, Proakis and Hansen 2000: 102)

The most important source of all sounds is the larynx, which is situated at the upper end of the trachea, an air tube stemming from the lungs (Rossing 1983: 282). The larynx is formed of two movable and elastic vocal folds and the narrow space between them is referred to as the glottis. The ends of the vocal folds are attached to the plates in the trachea with which the size of the slit between the folds can be adjusted. The front ends of the folds are attached to the thyroid cartilage whereas the back ends are attached to the arytenoids cartilages, which are moved by small muscles. The slit is at its largest during normal respiration and is almost closed during the production of sounds.

The diaphragm is used to control how the air is pressed from the lungs through the trachea and eventually through the slit in the larynx to create a sound. When the air is forced through the larynx, the vocal folds start to vibrate. With males the fundamental frequency F0 of this vibration is approximately 80-150 Hz and with females the corresponding frequency is approximately 150-300 Hz (Wiik 1998: 18). With children the frequency is usually 300 Hz.

2.2 Formants

The way the created airwaves that are emitted from the larynx will actually sound depends heavily upon the shape of the tract that the waves enter. The waves emitted from the larynx first permeate through the pharynx and then enter the oral cavity and the nasal cavity. The way the articulation will sound depends also on which cavity is open and which is obstructed. In any case, this male acoustic tube, which is approximately 17,5 centimeters in length and which consists of the larynx, the pharynx and the oral cavity is referred to as the vocal tract (Rossing 1983: 286). The female vocal tract is generally shorter and with children the cavity is the shortest, which means that children exhibit the highest formant patterns (Dew and Jensen 1977: 204). The other tract, which is separated by the soft palate (velum) and which enters the nasal cavity is referred to as the nasal tract.

As stated above, the vocal tract and the nasal tract affect the pulses stemming from the glottis and in this sense the vocal tract is an acoustic filter. This is essentially the source-filter theory introduced by Gunnar Fant (1960: 15). The shapes, the contours and the profile of the vocal tract determine the characteristics of this filter, which generates several moving resonance spots, which are referred to as formants. They manifest themselves as envelopes that modify the various harmonics of the source sound and each formant corresponds to one or several resonances in the vocal tract (Rossing 1983: 289). Rossing also points out that the formant frequencies are virtually independent of the source spectrum. According to Rosner and Pickering (1994: 4), human vocal tracts are able to express between 5 to 7 different formants during vowel production. They are essentially the poles of a transfer function, which may also entail zeros. The resonances in a tube, which is open at the other end and closed at the other, can be calculated according to the following equation:

$$F = \frac{n \cdot c}{4 \cdot l} \tag{1}$$

where n is an odd number, $c = 35000\frac{cm}{s}$ is the speed of sound at a temperature of approximately $35C^{\circ}$, which is the approximate temperature in the vocal tract and $l = 17{,}5cm$, which is the approximate length of the male vocal tract. With Equation 1, it is possible to calculate the formants in a tube, where the diameter is constant. The sound of such a tube is referred to as a neutral vowel (Wiik 1980: 15). The formants would have to be calculated using odd numbers for n, due to the fact that the tube is open at the other end and there is no node in the sound wave at the exiting end. With n=1,3,5,7, the formant values are the following:

F1 = 500 Hz

F2 = 1500 Hz

F3 = 2500 Hz

F4 = 3500 Hz.

The female vocal tract is generally shorter and therefore the female formant frequencies will be higher than those of males (Rosner and Pickering 1994: 15). In fact, T. Chiba and M. Kaijiyama stated as early as 1941 that the female vocal tract is approximately 13 % smaller than the male vocal tract with regard to the dimensions (1958: 189). However, further studies have shown that the “size difference is not always constant across all dimensions.” (Rosner and Pickering 1994: 57).

The shape of the vocal tract, on the other hand, can be adjusted most prominently with the tongue and its movements. By changing the width of the base of the tongue, the size of the surface area of the pharynx can be adjusted and by changing the position of the tip of the tongue, the tract can be narrowed or closed altogether. The width of the base of the tongue affects the back vowels whereas the position of the tip of the tongue affects some of the consonants such as the lateral /l/, the tremulant /r/ and the unvoiced alveolar plosive /t/. It is important to notice that the jaw and the lips also contribute to the shape and the acoustic features of the vocal tract. However, the entire vocal tract is a complex system, which is divided between many parameters (the tongue, the lips, the jaw etc.). These parameters are continuous in nature and divided over the entire system. Changing one of these parameters alone will often not affect the parameters of the transfer function of the tract. However, changing a combination of the parameters (the tongue, the lips etc.) will produce a certain transfer function.

2.3 Vowels and consonants and the F1/F2 acoustic plane

The different sounds in any language can be divided into two basic groups: the vowels and the consonants (Wiik 1998: 35). The sounds that humans produce can be either voiced or unvoiced. The vowels, which are at the center of focus in this study, are voiced (except in a situation where the vowel is whispered). With vowels, the nasal tract in closed by the velum and the air travels from the glottis through the vocal tract formed by the shape and position of the tongue and the shape of the lips. In some cases, vowels may also become nasalized, when the velum is lowered and some of the air enters the nasal cavity.

The vowels are relatively open when they are voiced and the shape and the different parts of the vocal tract remain relatively stable during the voicing of the vowels. The Finnish sound system consists of eight vowels (ɑ, e, i, o, u, y, œ and ɵ) and sixteen consonants (d, h, j, k, l, m, n, ng, p, r, s, t, v) along with b, f, and g from other languages. The vowels have four different characteristics that define them. A vowel can be a front vowel or a back vowel depending upon the location/position of the tongue. It was Daniel Jones who first suggested the use of a two-dimensional graph to illustrate the main features of vowel production (Rosner and Pickering 1994: 11). This two-dimensional graph has also been referred to as the vowel quadrilateral, vowel diagram or vowel chart. In it, the first two formants, the center frequencies F1 and F2, determine the vowel in question. Rosner and Pickering argue that the first two formants determine the vowel categorization or the perceived vowel quality (1994: 13). Rossing states that two or three formants are enough to identify a vowel (1983: 301). However, Gunnar Fant points out that a detailed analysis of the specifications of a vowel includes the frequencies, bandwidths and amplitudes of three or four formants (2004: 202). Within the scope of this study regarding the Finland Swedish back vowel /ʉ/, the analysis of the third and the fourth formant will be excluded since F3 has a negligible perceptual role for back vowels (Boe L-J, Schwartz J-L. and Vallee N. 1994: 195).

The Swedish front vowels (i:, y:, I, Y, e:, etc.) can be seen in Figure 2 at the left side of the vowel diagram. The back vowels are at the right side of the diagram (u:, o:, a: etc.).

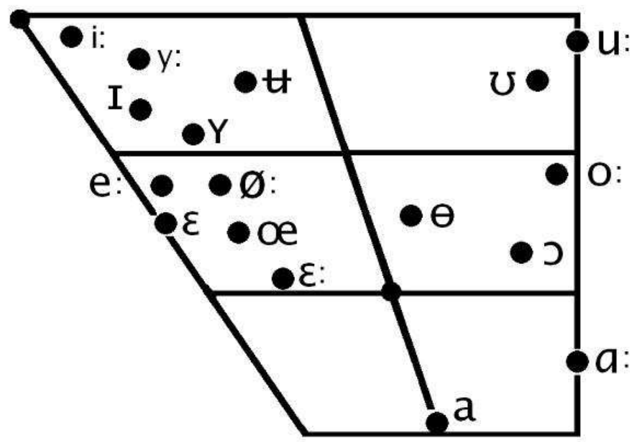


Figure 2 The Swedish vowels (modified from original) (Engstrand 1999: 140)

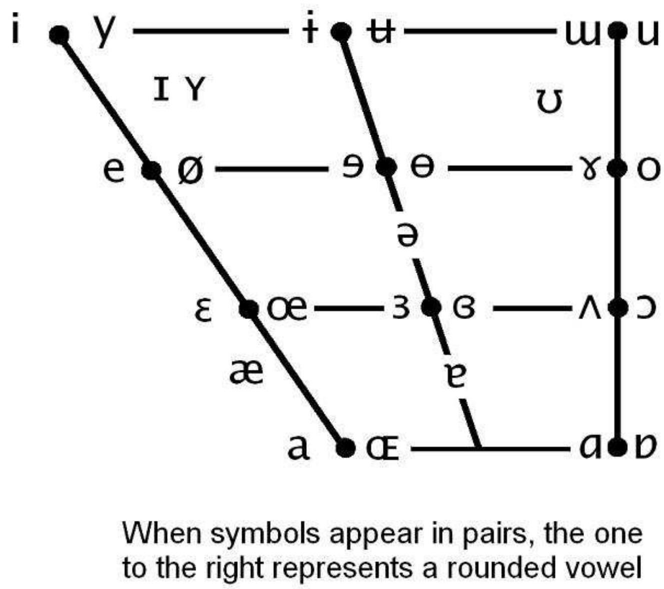


Figure 3 IPA representation of vowels (modified) (International Phonetic Association 2007)

Figure 3 demonstrates the phonetic representation of vowels in the vowel diagram according to the International Phonetic Association. The vowel can be an open or a closed vowel depending upon the size of the constriction between the palate and the tongue. The size of the constriction also contributes to the height of the vowel: whether the vowel is a high vowel or a low vowel. In other words, the tongue can be positioned either high or low in the mouth and the aperture of the jaw can be adjusted as well. The different levels in height are the following:

1) close vowel (high vowel)
2) near-close vowel
3) close-mid vowel
4) mid vowel
5) open-mid vowel
6) near-open vowel
7) open vowel (low vowel)

The roundedness of the vowel is the final characteristic, which provides information on whether the lips are rounded or not.

Table 1 illustrates the different characteristics of the 8 Finnish vowels.

Table 1 The Description of Finnish vowels

VOWELS		Front vowels		Back vowels	
	Height:	Wide	Round	Wide	Round
Close vowels	High	i	y		u
Mid vowels	Mid	e	ö		o
Open vowel	Low	ä		a	

2.4 Articulatory phonetics

Phonetics is a science, which investigates speech and how the different elements of speech, namely vowels and consonants, are formed. It also examines their acoustic nature and how the different sounds are perceived. Phonetics can be divided into three subcategories: 1) articulatory phonetics, 2) acoustic phonetics and 3) auditory phonetics (Wiik 1998: 11). Acoustic phonetics deals primarily with soundwaves and the analysis of the acoustic waveform whereas auditory phonetics studies mainly the function of the ear and the perceptual response to speech sounds e.g. in listener trials (Deller, Proakis and Hansen 2000: 116). Articulatory phonetics examines and studies how the different sounds of speech are formed and produced with the vocal tract and how the different articulators of the vocal tract (the tongue, the lips, the jaw, the palate, the teeth, etc.) are related to the process. In essence, articulatory phonetics deals with the sounds of speech. Speech production, on the other hand, can be described as a sequence of asynchronous movements from one set of articulatory targets to the next (Perkell and Nelson 1987: 187).

What is astounding is the fact that the flexibility and the almost indefinite amount of different expressions in a language are based upon a very limited amount of units that discriminate and separate different meanings. These units are referred to as phonemes. A phoneme is an abstract concept in the sense that even though a certain phoneme such as /ʉ/ may be perceived as /ʉ/ every time, the actual acoustic production of this phoneme may vary according to the context. In other words, it may depend upon the individual speaker or the context of the word, sentence, syllable or sound. Two speakers may produce an /ʉ/, which are both perceived as /ʉ/ by the listening audience, but which may still vary acoustically when measured with a suitable tool/equipment such as The Vowel Game. Rosner and Pickering state that vowel perception entails two processes. One is the categorization of different vowels and the second is the identification of the same vowel under different circumstances (1994: 1). While a phoneme is an abstract unit, which can se seen as a group of tangible sounds that all have the same linguistic function, it also important to establish the concept of the allophone. Two sounds that belong to the same phoneme are allophones. In other words, they “represent slight acoustic variations of the basic unit” (Deller, Proakis and Hansen 2000: 116).

The reason why persons who are skilled in a language are able to hear these abstract phoneme units is the fact that the phonemes carry a clear linguistic function. It has been stated that while the speech signal is a discreet queue of minimalistic units that define the linguistic differences in speech, it is also an acoustic continuous signal where the transitions from one phoneme to another are relatively smooth.

2.5 The Categorization of speech sounds and prototypes

The term categorical perception refers to the ability of human beings or animals to perceive and divide different sensory phenomena into different categories according to certain specific criteria. With regard to speech, categorical perception allows the hearer to distinguish two sounds from one another and place them into their respective categories. Obviously, this type of perception can be applied and broadened to other types of phenomena as well, but within the confines of this study, categorical perception refers to the categorization of speech sounds, namely vowel sounds.

In the late 50s, Alvin Liberman claimed that categorical perception was unique to speech. In his theory, categorical perception referred to a phenomenon where a perceived quality of a phoneme changes suddenly from one category to another and that the origin of this theory can be traced to the anatomy of speech production. This was essentially the motor theory of speech. Liberman's old form of the motor theory is from 1967 and he revised it in 1985 with I.G. Mattingly (1985: 1). Carol Fowler, on the other hand, offered the direct realism theory in which the objects of perception are the actual vocal tract gestures as opposed to the abstract phonemes/sounds presented in Liberman's motor theory (Fowler 1977). Carol Fowler's direct realism theory and Liberman's motor theory of speech form strong articulatory theories. In essence, they center on the neuromotor commands and articulatory gestures and in these theories the listener uses the acoustic products of articulation to gain access to the speaker's neuromotor commands or articulatory gestures (Rosner and Pickering 1994: 372).

The next question within this study is, of course, to determine how vowel sounds are perceived and categorized. As early as 1952, G. E. Peterson studied vowel imitation as a plausible method for defining categories for vowels. In his studies, the test subjects imitated their own productions of vowels in /CVC/-syllables (Consonant-Vowel-Consonant). The results showed that the formant frequencies of the imitations remained close to the original productions thus suggesting categorization (Rosner and Pickering 1994: 114). Later in 1966 L.A. Chistovich explored subjects who imitated synthetic vowel stimuli and again the results indicated clear proof of categorical production. Categorical production was further studied by R.D. Kent, B.H. Repp and D.R. Williams and they all came to similar conclusions. Still, the tests involving imitation did not answer the question of whether or not the categorical tendencies in production relied on the perceptual categories for vowels (ibid.).

In her studies in 1991, Patricia K. Kuhl stated that “Many perceptual categories exhibit internal structure in which category prototypes play an important role” (1991: 93).
Her studies, which explored the internal structure of phonetic categories, involved adults, infants and monkeys and one of the key findings was that within a vowel category, there was a certain location, which the human listeners rated as the best instance or prototype of the vowel. When the stimuli were removed further away from this prototype, the perceived “goodness” of the vowel declined (ibid.). The concept of the perceptual magnet effect was introduced when the effects of these prototypes were examined on speech perception. Fundamentally, the term refers to a phenomenon within the vowel category where the neighboring vowels of the prototype start to assimilate into the prototype. In this concept, the listener perceives the neighboring vowels as one and the same, unlike the vowels, which are further away from the prototype.

In the testing of The Vowel Game, the prototype and the audio model for the prototype of /ʉ/ was selected on the basis of the Turku Vowel Test, which was initialised on the Internet in 2000 (Raimo, Savela and Aaltonen 2002: 45). In the test, subjects were requested to evaluate the goodness of 400 synthetic vowel stimuli according to their native tongue. 16 subjects took part in the Finland Swedish segment, where they determined, among other prototypes, the prototype of the Finland Swedish /ʉ/-category. According to the tests, the prototype had the following formant values:

$$F1 = 300 \text{ Hz}$$
$$F2 = 969 \text{ Hz}$$

The synthetic prototype was produced using a synthesizer, KLATT, and the F1 values of the produced stimuli varied between 250-800 Hz, whereas the F2 values varied between 600-2800Hz.

The vowel categories defined in The Vowel Game were formed according to the data gathered from the tests of the Turku Vowel Test. A number of individuals, who took part in the Turku Vowel Test, evaluated 400 synthetic stimuli on a scale of 1-7 and thereby created the categories and the average prototypes for each of the vowel categories present in The Vowel Game. The Finnish vowel diagram, as defined in the Turku Vowel Test, is illustrated in Figure 4. This is the vowel diagram also used in The Vowel Game.

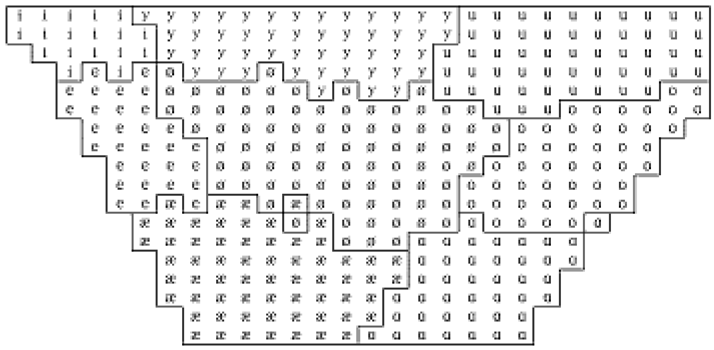


Figure 4 The Finnish vowel diagram (Raimo, Savela and Aaltonen 2002: 49)

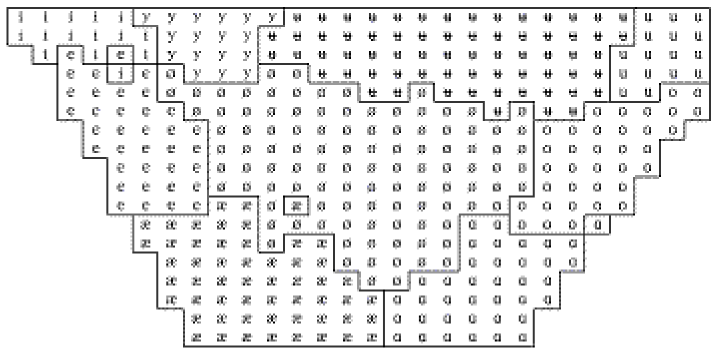


Figure 5 The Finland Swedish vowel diagram (Raimo, Savela and Aaltonen 2002: 49)

Figure 5 illustrates the Finland Swedish vowel diagram, as defined in The Turku Vowel Test. The prototype of the /ʉ/-category, according to the 16 subjects who took the Finland Swedish test segment, is located on the 3rd row from the top and is the 8th vowel from the right. The prototype can be seen circled in Figure 6. The box appears lighter than the other vowels in the category. This indicates that the vowel was deemed the best example of the Finland Swedish /ʉ/-vowel according to the 16 subjects.

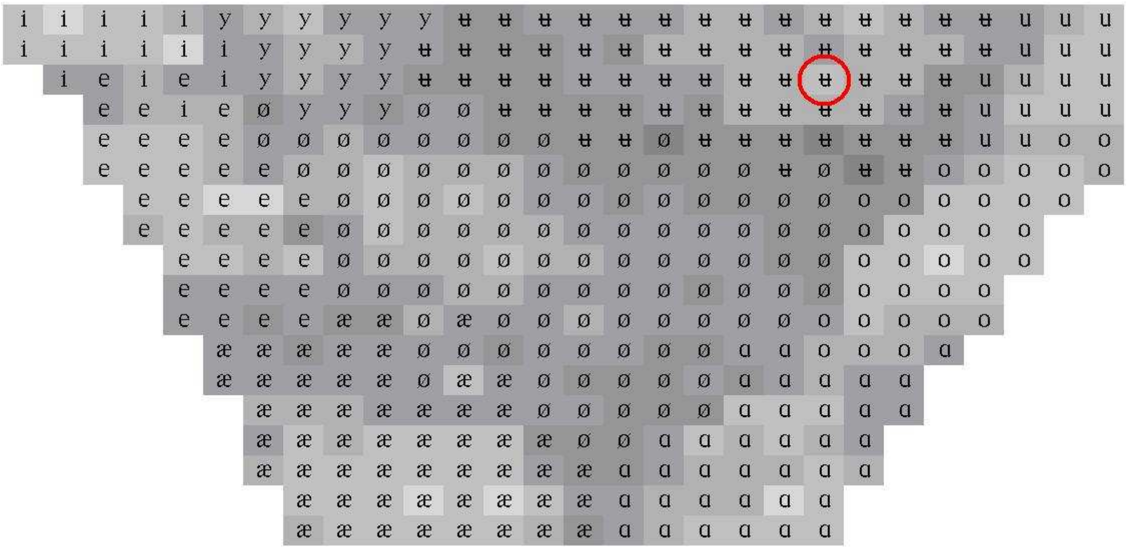


Figure 6 The goodness of the Finland Swedish vowels (Raimo, Savela and Aaltonen 2002)

The natural speech audio model for /hʉs/ was obtained from the practise audio tape of the *Va va de du sa: Gröna boken* by Pedersen, Rosenberg-Wolff and Uddström (1996a).

The F1/F2-plane and formants, which were introduced in Chapters 2.2 and 2.3 represent vowels as physical phenomena or stimuli. Still, Rosner and Pickering argue that they do not represent vowels as auditory events. Furthermore, Rosner and Pickering also assess that the F1/F2-plane is not a satisfactory model in itself for the theory of vowel perception. An auditory transform must be performed on the axes of the F1/F2-plane so that the frequency variable changes into an auditory variable with an identifiable pitch or auditory filter rate (1994: 115). The most common examples of such auditory transforms are, for instance, log, Koenig, mel, Bark and ERB (Equivalent Rectangular Bandwidth). In essence, they are psycho-acoustic scales, of which mel is probably the most prominent. The mel-scale was derived from experiments by Stevens and Volkman in 1940 (Rosner and Pickering 1994: 17). The linear frequency scale of Hertzes can be converted to the mel-scale using Equation 2.

$$mel = 1127{,}01048 * \ln(1 + \frac{f}{700}) \qquad (2)$$

The vowel boxes in The Vowel Game’s diagram appear to be identical in size, but the actual values of the boxes are expressed using the mel-scale whilst keeping in mind that the results in Chapters 6 and 7 are analysed using the Hertz-scale. 1000 Hz corresponds to 1000 mel according to Equation 2. All the frequencies above 1000 Hz, correspond to progressively smaller values on the mel-scale. This means that if the boxes in the diagram were to be adjusted according to the frequency scale, all the boxes in the horizontal F2 plane would be wider above 1000 Hz due to the fact that the differences in pitch above that level are perceived differently from the frequencies below 1000 Hz. As an example, the psycho-acoustic difference between 500 Hz and 600 Hz is 90,2 mels whereas the difference between 1500 Hz and 1600 Hz is only 50,1 mels according to Equation 2. One mel corresponds to the smallest perceivable psycho-acoustic difference between the pitch of two fundamental frequencies. In this sense, the listener's perceived difference between 500 Hz and 600 Hz can be defined more precisely than the difference between 1500 Hz and 1600 Hz, even though the difference is the same on the Hertz scale (100 Hz) in both cases.

2.6 The Effects of age on native-like pronunciation

All The Vowel Game’s test subjects were adults. Therefore, one of the questions before the tests was whether or not it was feasible to teach pronunciation to adults. According to a popular belief, children are more proficient at learning languages than adults. It has been suggested that the earlier the learning starts, the better. Another common claim is that the learning will be considerably more strenuous at a later stage, if the benefits of the young age are not utilised to their full potential. For instance, the communicative acquisitionist naturalistic (CAN) 'megatheory' claims that preschool children or children on the whole are more proficient in SLA (Second language learning) than young adults (Hammerly 1991: 8). Over the years, this view has emerged in several examples all over the world. For instance, in 1959 Noah Chomsky described how the parents of an immigrant child became frustrated over the fact that their child was able to master the different subtleties in the foreign language in a relatively short period of time whereas the parents had to struggle with the language all their lives with constant effort (Chomsky in Cook 1991: 83).

These views on young age were summoned in the Critical Period Hypothesis by E. H. Lenneberg, which states that human beings are only capable to learn languages between the age of two and the early teens. This view was seen to apply both for L1 (mother tongue) and L2 (second language) learning (Lenneberg 1967: 176). The reason for adults' losing their ability to learn was suggested by several theories. One of the theories covered the physical aspects of the brain, namely the loss of plasticity in the brain and "the lateralization of the language function in the left hemisphere in the brain" (Ellis 1985: 107). Another theory dealt with social factors of how children encounter certain situations and relationships in a different way from adults. The third theory concerned the cognitive explanations of how the learning of a natural language was interfered by the abstract way of thinking, characteristic of adults (Cook 1991: 83). Therefore, making sure that the child starts learning a second language as early on as possible is pivotal.

What is startling, however, is that the evidence of the child superiority in research is hard to find (Cook 1991: 84). Some research indicates that in certain learning environments, adults actually fare better. For instance, Asher's and Price's research in 1967 showed that when the Total Physical Response teaching method was applied in teaching Russian to both children and adults, the older students outperformed the younger (ibid.). Similar results were received when an immersion technique was applied in Canada where English-speaking students were taught the curriculum through French. Again, the late immersion pupils fared better. It has been concluded that children learn a second language better in an environment that is more suited and natural to them as opposed to the typical adult learning environments. When children and adults are taught the same way, adults generally have the advantage over the children (ibid.).

However, the most prominent advantage that children may have over adults lies in pronunciation. It has been claimed that proper pronunciation and convincing accent cannot be achieved if the learning takes place after the early teens. For instance, Hector Hammerly claims that young children are better in "untutored acquisition of native-like pronunciation and intonation from natives in the environment" (1991: 8). Still, various researches have been conducted on this and the results have been somewhat contradictory. The researches of e.g. Asher and Garcia in 1969 and Ramsay and Wright in 1974 showed that younger children had the advantage over older children with regard to pronunciation. These findings, on the other hand, were challenged by researches done by Cummins in 1981 and by Snow and Hoefnagel-Hohle who proved in 1977 that Dutch imitation by English-speaking students was more proficient with older students (Cook 1991: 84).

Today the short term benefits of youth are usually examined separately from the long term disadvantages of age. (Cook 1991: 85) The one conclusion that John Singleton made in 1989 supports the initial assumption regarding age:

> "The one interpretation of the evidence which does not appear to run into contradictory data is that in naturalistic situations those whose exposure to a second language begins in childhood in general eventually surpasses those whose exposure begins in adulthood, even though the latter usually show some initial advantage over the former."

Still, with these types of statements, one should always take into account the amount of language exposure and interaction and remember that in general learning adults are better with regard to cognitive maturity, learning strategies and study habits (Hammerly 1991: 8). However, as far as the native-like pronunciation is concerned, the young age in learning does bring a distinctive advantage.

In the tests of The Vowel Game, all the test subjects were adults with an average age of 22,35 years (see Table 6) and the primary goal was not to teach the subjects native-like pronunciation but rather to investigate if The Vowel Game has an effect in making their pronunciation of the /ʉ/-vowel more constant. The evidence of the research provided in this chapter reinforces the assumption that the teaching of native-like pronunciation to adults might prove to be problematic. As Dowd, Smith and Wolfe point out, adults who are learning foreign languages rarely acquire an authentic pronunciation (1998: 1).

3. The Mathematical methods in The Vowel Game

This chapter presents the various mathematical methods, which are used in the process of retrieving formants from the speech signal. First, Chapter 3.1 centers on the vocal tract and how speech production can be modelled in general. Chapter 3.2 takes a closer look at the autocorrelation method, which was the linear predictive analysis technique used in The Vowel Game. Chapter 3 centers on the method with which the 10^{th} order LP coefficients are solved in The Vowel Game and finally Chapter 3.4 explains the fundamentals of the Fourier series and the Fourier transform, which is performed on the impulse response of the LP coefficients to obtain the spectral envelope of the speech signal. It is from this spectral envelope that the formants of the speech are received.

3.1 The Principals of formant retrieval

The Vowel Game utilises 10^{th} order linear predictive analysis to retrieve the first two formants from the speech signal, which are crucial in determining the vowel in question. According to Rabiner and Schafer, linear predictive analysis is one of the most powerful speech analysis techniques in use. With this method, it is possible to estimate several fundamental speech parameters such as F0, formants, spectra and the vocal tract functions with astounding accuracy. Linear predictive analysis is essentially focused on estimating a speech sample as a linear combination of previous speech samples (1978: 396). In it, the sum of squared differences between the speech samples and the estimated samples (which have been acquired via linear prediction) is minimized, thereby creating a collection of prediction coefficients.

According to Rabiner and Schafer, there have been at least 7 different formulations of the linear predictive analysis (1978: 397).

1) The Covariance method
2) The Autocorrelation formulation
3) The Lattice method
4) The Inverse filter formulation
5) The Spectral estimation formulation
6) The Maximum likelihood formulation
7) The Inner product formulation

The linear predictive analysis method chosen for The Vowel Game is the autocorrelation formulation. Figure 7 illustrates the fundamental principle of how speech production can be modelled.

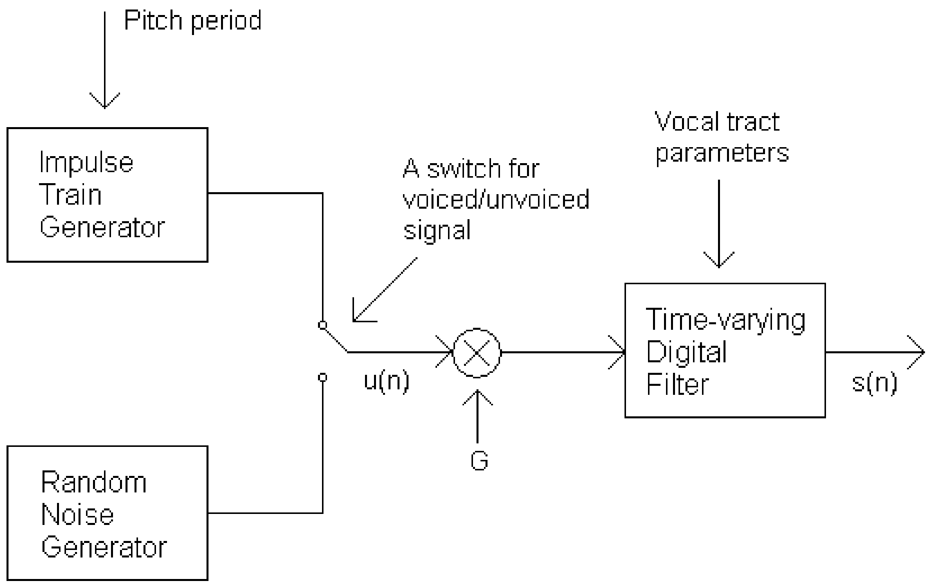


Figure 7 A model for speech production (modified) (Rabiner and Schafer 1978: 398)

The time-varying digital filter in Figure 7 can be represented by a function, which is fundamentally a composite of spectrum effects of radiation, vocal tract and glottal excitation (Rabiner and Schafer 1978: 398). The function is of the form:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} \qquad (3)$$

The parameters in the above formula are the following: S(z) is the output signal and U(z) is the input signal in the frequency domain. G is the gain parameter and [a_k] are for the coefficients of the digital filter. The above function is an all-pole model, which can be used with non-nasal voiced sounds. This is beneficial due to the fact that the sounds, which are examined in this study, are vowels and vowels are prominently voiced and non-nasal.

The speech samples in the time domain can be defined according to the following equation:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \qquad (4)$$

The above equation basically describes how the speech samples are related to the excitation. The linear predictor, which carries the linear predictor coefficient α_k is defined as:

$$\bar{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \tag{5}$$

The system function of a linear predictor, which is of the p^{th} order, is presented in the frequency domain as follows:

$$P(z) = \sum_{k=1}^{p} \alpha_k z^{-k} \tag{6}$$

It is essential to define the prediction error, which can be expressed as a difference between the speech sample and the linear predictor:

$$e(n) = s(n) - \bar{s}(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) \tag{7}$$

Equation 7 can be simplified in the following manner, assuming that
$a_k = \alpha_k$. So if $a_k = \alpha_k$, then:

$$e(n) = \sum_{k=1}^{p} \alpha_k s(n-k) + Gu(n) - \sum_{k=1}^{p} \alpha_k s(n-k) = Gu(n) \tag{8}$$

If the p^{th} order system function of the linear predictor is P(z) as defined in Equation 6, the transfer function of the inverse filter is defined as:

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} \tag{9}$$

Therefore, the original definition of the time-varying digital filter, as described in Equation 3, can now be expressed as:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{G}{A(z)} \tag{10}$$

The predictor coefficients must be determined from short segments of speech due to the fact that the speech signal varies according to time. Therefore, the goal is to determine a group of predictor coefficients, which minimize the mean-squared prediction error over a short segment of speech. These coefficients can be deemed or assumed to be the parameters of the system function H(z) (Rabiner and Schafer 1978: 399).

Rabiner and Schafer state that since it was established that if $a_k = \alpha_k$, then $e(n) = Gu(n)$. This would indicate that the prediction error $e(n)$ is composed of a train of impulses and this, on the other hand, would suggest that the prediction error would be prominently miniscule. Rabiner and Schafer also state that the fundamental reason for using the minimum mean-squared prediction error is the fact that the method produces a collection of efficiently solvable linear equations that in turn produce the predictor coefficients. This method offers a very precise presentation of the speech signal (1978: 400).

Since the speech signal is constantly changing with respect to time, the short-time average prediction error must be defined in the following manner:

$$E_n = \sum_m e_n^2(m) = \sum_m (s_n(m) - \bar{s}_n(m))^2 = \sum_m \left[s_n(m) - \sum_{k=1}^{p} \alpha_k s_n(m-k) \right]^2 \qquad (11)$$

In the above equation, $s_n(m)$ represents a speech segment, which has been chosen from the close proximity of the n^{th} sample. ($s_n(m) = s(m+n)$).

To find values for α_k that minimize the short-time average prediction error E_n, the following partial derivative is set to zero:

$$\frac{\partial E_n}{\partial \alpha_i} = 0 \qquad i = 1,2,..., p \qquad (12)$$

This, in turn, produces the following equations, where $\hat{\alpha}_k$ depicts the values of α_k that minimize E_n:

$$\sum_m s_n(m-i)s_n(m) = \sum_{k=1}^{p} \hat{\alpha}_k \sum_m s_n(m-i)s_n(m-k) \qquad 1 \le i \le p \qquad (13)$$

$\sum_m s_n(m-i)s_n(m-k)$ can be expressed as:

$$\phi_n(i,k) = \sum_m s_n(m-i)s_n(m-k) \qquad (14)$$

Therefore, Equation 13 can be expressed as:

$$\sum_{k=1}^{p} \alpha_k \phi_n(i,k) = \phi_n(i,0) \quad i = 1,2,..., p \tag{15}$$

The above equation is the definition of p equations, with p unknown prediction coefficients α_k that minimize the average squared prediction error for the segment $s_n(m)$ (ibid.).

According to Rabiner and Schafer, the minimum mean-squared error can be expressed in the following manner. The form has been obtained by using Equation 11 and Equation 13 (1978: 401).

$$E_n = \sum_m s_n^2(m) - \sum_{k=1}^{p} \alpha_k \sum_m s_n(m) s_n(m-k) \tag{16}$$

Equation 16, which depicts the short-time average error, can be expressed in a more compact form by using the earlier definition described in Equation 15:

$$E_n = \phi_n(0,0) - \sum_{k=1}^{p} \alpha_k \phi_n(0,k) \tag{17}$$

As Rabiner and Schafer explain, the total minimum error is comprised of a constant component and a component, which is dependent of the predictor coefficients (α_k) (ibid.).

The next step is to solve the predictor coefficients. For that, the quantities $\phi_n(i,k)$ for $1 \le i \le p$ and $1 \le k \le p$ must be solved. After that, the predictor coefficients can be determined by solving Equation 15.

Rabiner and Schafer state that the limits for the sums (eg. $1 \le i \le p$ and $1 \le k \le p$) must be over a finite interval in order to achieve a short-time analysis procedure. This leads to the linear predictive analysis method chosen for The Vowel Game, which is the autocorrelation formulation.

3.2 The Autocorrelation method

This chapter centers on the autocorrelation method in which the audio signal is segmented using a finite length window referred to in The Vowel Game's case as a Hann-window. However, in order for us to perform any operations on the signal, the signal must first be obtained from the microphone line and then converted to digital form by using an analog to digital transform. The microphone first delivers the signal in the form of a changing voltage. This changing voltage is consequently sampled using a suitable sampling frequency. In The Vowel Game, the sampling frequency has been set according to the application's need. There are no significant elements in speech with regard to vowels above 4 kHz and therefore the sampling frequency has been set according to Nyquist's theory at 8 kHz to avoid aliasing (Carlson 1986: 353). Nyquist's theory is presented in Equation 18.

$$f_s = 2 f_1 \qquad (18)$$

On a side note, O'Shaughnessy points out that the most relevant information in speech with regard to communication is in the range of 200-5600 Hz (1999: 109). After sampling, the signal is quantized using a 8-bit quantizition. This provides $2^8 = 256$ different levels for the samples. It is only after these operations that a segment can be extracted from the waveform for windowing. The signal must also be pre-emphasized by 6dB per octave after the windowing. This means that the energy of the high frequency spectrum is increased relatively to the lower frequency spectrum to ensure that both spectrums are weighted equally (Paganus et al. 2006: 700).

When a segment, $s_n(m)$, is taken out from the speech waveform, it is assumed that the waveform segment is zero outside of its interval $0 \le m \le N-1$ (Rabiner and Schafer 1978: 401). In the case of The Vowel Game, a finite length window called Hann-window was used, which is identically zero outside the interval stated above. The length of the window in the application is 32 ms. The notation of the Hann-window has been set as $w(m)$ and so the waveform segment can be expressed as:

$$s_n(m) = s(m+n)w(m) \qquad (19)$$

Therefore, if $s_n(m)$ is non-zero for the interval $0 \le m \le N-1$, the prediction error $e_n(m)$ for the p^{th} order predictor will be non-zero for the interval $0 \le m \le N-1+p$ (ibid.). As stated earlier, The Vowel Game utilises 10th order linear prediction.

So, in general, a waveform segment is formed by multiplying the signal with a window coefficient (e.g. a Hann-window) as in Equation 19. The Hann-window $w(m)$ can be expressed in the following manner:

$$w(m) = \frac{1}{2}\left[1-\cos\left(\frac{2\pi m}{N-1}\right)\right] \qquad 0 \le m \le N-1 \qquad (20)$$

The prediction error, E_n, can therefore be expressed as a sum with the following limits:

$$E_n = \sum_{m=0}^{N+p-1} e_n^2(m) \qquad (21)$$

Rabiner and Schafer also point out that the prediction error could have been expressed as a sum over all nonzero values from $-\infty$ to ∞ (ibid.) In addition to this, they observe that the probability that the prediction error is large at the beginning of the interval ($0 \le m \le 1-p$) is high due to the fact that the samples of the signal, from which the prediction is made, have been set to zero. Similarly, the probability that the error is large at the end of the interval ($N \le m \le N+p-1$) is also high because the samples, from which the prediction is made, are nonzero.

Earlier it was defined that:

$$\phi_n(i,k) = \sum_m s_n(m-i)s_n(m-k)$$

$$\phi_n(i,k) = \sum_{m=0}^{N+p-1} s_n(m-i)s_n(m-k) \qquad 1 \le i \le p \qquad (22)$$

$$0 \le k \le p$$

This can be expressed as:

$$\phi_n(i,k) = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k) \qquad 1 \le i \le p \qquad (23)$$

$$0 \le k \le p$$

The above form is identical to the short-time autocorrelation function, which has been evaluated for ($i-k$) (Rabiner and Schafer 1978: 402).

$$\phi_n(i,k) = R_n(i-k) \tag{24}$$

$R_n(k)$, which is an even function, can therefore be expressed as:

$$R_n(k) = \sum_{m=0}^{N-1-k} s_n(m) s_n(m+k) \tag{25}$$

This is because:

$$\phi_n(i,k) = R_n(|i-k|) \qquad i = 1,2,\ldots,p \qquad k = 0,1,\ldots,p \tag{26}$$

The earlier definition (Equation 15):

$\sum_{k=1}^{p} \alpha_k \phi_n(i,k) = \phi_n(i,0)$ can now be expressed as:

$$\sum_{k=1}^{p} \alpha_k R_n(|i-k|) = R_n(i) \qquad 1 \le i \le p \tag{27}$$

Equation 27 finally leads to the matrix form below:

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & \ldots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \ldots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \ldots & R_n(p-3) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ R_n(p-1) & R_n(p-1) & R_n(p-1) & \ldots & R_n(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \ldots \\ \ldots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \ldots \\ \ldots \\ R_n(p) \end{bmatrix} \tag{28}$$

The above matrix is a pxp Toepliz matrix, meaning it is symmetric and all the elements in a certain diagonal are equal (Rabiner and Schafer 1978: 402). In The Vowel Game's case, the matrix is a 10x10 Toepliz Matrix, since the application utilises the 10^{th} order linear prediction. The next step is, of course, to find a way of solving the predictor coefficients. According to Rabiner and Schafer, Durbin's recursive solution is the most effective (1978: 411).

3.3 The Solution of the predictor coefficients

The equation for the matrix in Chapter 2.7, which dealt with the autocorrelation method, was the following (defined earlier in Equation 27)

$$\sum_{k=1}^{p} \alpha_k R_n(|i-k|) = R_n(i) \; 1 \le i \le p \tag{27}$$

The matrix (Equation 28) is a Toepliz-matrix, as stated earlier, and to solve the coefficients of the matrix, there have been a few different popular methods in use over the years, e.g. the Levinson and Robinson method (Rabiner and Schafer 1978: 411). However, according to Rabiner and Schafer, the most efficient of the methods is Durbin's recursive solution. The principal of this method is to first solve e.g. a 3x3 matrix and then proceed to solve a 4x4 matrix by utilising the previous result from the 3x3 matrix until the entire matrix is solved (Paganus 2007: 29) and (Deller, Proakis and Hansen 1993: 299). However, there are also a number of other ways of solving the predictor coefficients. E.g. there is the Cholesky decomposition solution for the covariance method (Rabiner and Schafer 1978: 407). The coefficient matrix in it is not a Toeplitz-matrix and Durbin's recursive method cannot therefore be used. However, the decomposition solution used in the covariance method can also be used for the Toeplitz-matrix. It was therefore used in The Vowel Game as well. The decomposition method is referred to as LU and it is a slightly more general form of a method of solving the coefficients (Paganus 2007: 29). A matrix A, which is in the form:

$$Ax = b \tag{29}$$

can be decomposed into an Upper and a Lower matrix, hence the name for the decomposition solution: LU (Deller, Proakis and Hansen 2000: 309).

$$A = LU \tag{30}$$

Now, the x is solved using the following equations:

$$Ly = b \tag{31}$$

$$Ux = y \tag{32}$$

y can be solved by using Equation 31 and by using y and Equation 32, x can be solved as well.

Finally, the predictor coefficients can be presented in an LP polynome in the following manner (Paganus 2007: 31):

$$A(z) = 1 + \alpha_1 z^{-1} + ... + \alpha_p z^{-p} \tag{33}$$

Equation 33 is essentially the same as Equation 9. To finally retrieve the formants, we must determine the amplitude response of the following transfer function:

$$\frac{1}{A(z)} \tag{34}$$

The Maxima of the amplitude response can be deemed formants. First, the impulse response must be determined in the following manner (ibid.):

$$y(n) = \begin{cases} 1 & ,n = 0 \\ \alpha_1 y(n-1) + \alpha_2 y(n-2) + ... + \alpha_n y(0), & ,n > 0 \end{cases} \tag{35}$$

The impulse response $y(n)$ must be Fourier transformed. The result is the frequency response. Finally, to determine the amplitude response, the absolute value of the frequency response must be determined in the following way:

$$|FFT| = \sqrt{\alpha^2 + \beta^2} \tag{36}$$

Equation 36 represents the amplitude response where α is the real part of the frequency response and β is the imaginary part. From this we obtain a graph where the maximum peaks are deemed formants. The formant values are calculated using the following equation:

$$F_m = \frac{mf_s}{N} \tag{37}$$

where m represents the amplitude response values and N is the amount of samples in the impulse response (ibid.). The principles of the Fourier transform are explained in the following chapter.

3.4 The Fourier transform and the Fourier Series

In its standard form, the Fourier transform is an integral transform, which presents the signal's spectrum as a function of frequency. With the aid of the Fourier transform, it is possible to form a line spectrum, which is based upon the idea that the Fourier-series can be expressed as a sum of the direct current and the harmonic cosine waves.

When the formants of the signal cannot be detected from the signal in the time domain, it becomes necessary to perform the FFT (Fast Fourier Transform) of the signal in order to enter the frequency domain and perceive the actual components. The amplitude of the components becomes essential in the forthcoming analyses.

If the signal is denoted as $x(t)$, the Fourier transform is expressed according to Equation 38 (Carlson 1986: 32):

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt \qquad (38)$$

In The Vowel Game, the Fourier transform is performed on the impulse response defined in Equation 35. Basically, the Fourier transform presents the signal's spectral density and the used unit is either $\frac{V}{Hz}$ or $\frac{\sqrt{W}}{Hz}$. The Fourier transform is essentially a mathematical signal model, which cannot be measured using indefinite time signals. (This is obvious since the Fourier-transform is not a function of time by its definition).

The new variable f in Equation 38 is real, but the Fourier transform is often complex due to the fact that the signal $x(t)$ is multiplied by $e^{-j2\pi ft}$ where j is an imaginary unit. Its definition is the following:

$$j^2 = -1 \qquad (39)$$

The following expression can be used for the angular velocity:

$$\omega = 2\pi f \qquad (40)$$

Only the Fourier series of a periodical signal can be deemed to have a signal spectrum. The Fourier series is a function of time (as opposed to the Fourier transform which is a function of frequency). In general, the Fourier series are an orthogonal representation of a complex exponential base function. The base function is of the form:

$$\phi_n(t) = e^{j2\pi nt/T} \quad t \in \left[-\frac{T}{2}, \frac{T}{2}\right] \quad n = 0, \pm1, \pm2, \pm3... \tag{41}$$

The Fourier series of a signal $x(t)$ in the domain $t \in \left[-\frac{T}{2}, \frac{T}{2}\right]$ is expressed as follows:

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{\frac{j2\pi nt}{T}} \tag{42}$$

c_n is the Fourier coefficient, which can be expressed in the following way:

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t)\phi_n^*(t)dt = \frac{1}{T} \int_{-T/2}^{T/2} x(t)e^{-j2\pi nt/T} dt \tag{43}$$

In Equation 43, the integral is a correlation between the signal and the base function. The correlation has been divided by the length of the time period.

The Fourier series can be calculated for signals where the following conditions apply:

$$\int_{-T/2}^{T/2} |x(t)|^2 dt < \infty \tag{44}$$

$$\lim\{|x(t+\varepsilon) - x(t-\varepsilon)|\} \neq 0 \quad \text{(in a finite amount of points)} \tag{45}$$

The Equations/conditions 44 and 45 apply when the signal is limited with regard to its amplitude and has a bookable amount of discontinuities.

4. Vowel training systems

As of this date, The Vowel Game is the most recent product to emerge from the field of vowel training systems. It has been preceded by other systems, which have been developed, more or less, to train the student's pronunciation of vowels utilising visual real-time feedback and the vowel diagram presenting the F1/F2-plane. This chapter centers briefly on some of these earlier systems and their main differences with The Vowel Game. Systems such as Kay Sona-Match, Protrain, Dr. Speech, Video Voice and Accent Lab were also a part of Michael Carey's study in L1-specific CALL pedagogy for the instruction of pronunciation with Korean learners of English. In his study, Carey looked, among other things, at some of the most apparent disadvantages of the systems and argued that the systems utilising the F1-F2 plotting are potentially the most effective, although they present their own problems as well (2002: 161).

The studies of Anne-Marie Öster centered on CAPT (Computer-aided pronunciation training) and CBST (Computer-based speech therapy), which utilised visual feedback with focus on children with profound hearing impairments. In her study, she offers a number of criteria, which should be taken into account when visual feedback is used in speech therapy (2006: 197). She stresses that it is important to first diagnose the individual deviations that have the most effect on intelligibility. After that, the speech material must be planned and defined, which is followed by the instructions and the initial training phase. The final stage is the "repetitive training for generalisation, transfer and linguistic use" followed by an evaluation (Öster 2006: 198). The training systems presented in this chapter, SpeechViewer III and Trollerilådan, were also a part of Öster's studies and she offers her assessment of the former in this work.

Deborah Healey, among others, also explored a number of the aforementioned products and offered her constructive criticism related to their apparent inadequacies (1998). The point, which M. Carey brought up, for instance, was related to the debate regarding the relationship between the acoustic representation of a vowel with regard to F1 and F2 and the perception of it. Some, like Rosner and Pickering, argue that the F1/F2-representation is not an adequate model in itself for perception (see Chapter 2.5). In The Vowel Game, for instance, the categories are presented using the mel –scale.

4.1 The Vowel Game

The Vowel Game utilises a graphic vowel diagram with visible vowel categories that function in real time. The idea of The Vowel Game is what its title suggests. It is a game where the desired pronunciation brings the student points. The version, which was modified for the tests, excludes the game portion, but offers the user the possibility to record the pronunciation in text format. This way the formant values (Hz), along with the date and time and other optional information can be observed in a text file.

As Figure 8 illustrates, The Vowel Game has been modified for the sake of the tests. Buttons with clear labelling (in Finnish) have been added to make the actual test situation less complex and to ensure that all the necessary operations can be performed from The Vowel Game's interface. The three buttons at the lower left corner in the "Harjoittele"-section ("Practise"-section) in Figure 8 are listed as follows:

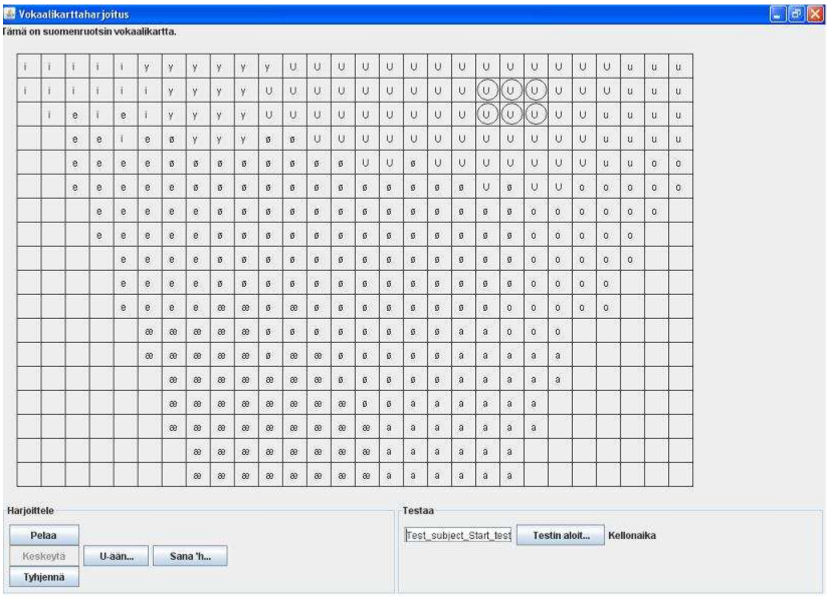


Figure 8 The Vowel Game

"Pelaa" (= Play)

This button activates The Vowel Game's vowel diagram so that all the sounds, which the microphone picks and which the program interprets as vowels, are plotted in the vowel diagram. 6 boxes have been circled in the /ʉ/-area in the diagram (the 2^{nd} and the 3^{rd} row from the top of the diagram and $7^{th,}$ 8^{th} and 9^{th}

column from the right. The circled box on the 3rd row and in the 8th column is the reference box. Its middle point is the reference point (F1 = 300 Hz, F2= 969 Hz). Whenever a hit is plotted in any of the 6 boxes, they turn red and remain red for the duration of the session. Whenever a hit is plotted again in any of the red boxes, they turn momentarily yellow. This is to ensure that all the consecutive hits in these circled boxes may be perceived.

“Keskeytä” (= Stop)

This button disconnects the microphone line so that The Vowel Game will not display any hits.

“Tyhjennä” (= Clear)

Thus button clears all the red boxes and turns their colour to the regular grey.

“U-ään…” (= The /ʉ/-audio model)

This button plays the synthetic /ʉ/-audio model through a set of headphones or speakers. This button was added to The Vowel Game for the sake of the tests.

“Sana ‘h…” (= The /hʉs/-audio model)

This button plays the natural /hʉs/-audio model through a set of headphones or speakers. This button was added to The Vowel Game for the sake of the tests.

The “Testaa” (=”Test”)-section includes one button and a text-field.

“Testin aloit…” (= “Start the test”)

This button plays the recording, used in all the tests. The recording lasts 77 seconds and first features the synthetic /ʉ/-audio model 10 times, with each model 3 seconds apart. After the 10 /ʉ/-audio models, the recording plays back the 10 natural /hʉs/-audio models, each 3 seconds apart. When this button is pressed, the playback of the audio recording is commenced along with the recording of the hits in the text file.

The test session and the test subject's name can be entered in the text field before pressing the **"Start the test"-**button. This way the text field's information is recorded in the text file, which helps in the identification of the data later on.

Figure 9 illustrates three hits in The Vowel Game's vowel diagram. Two are in the /ʉ/-area and one is in the /i/-area.

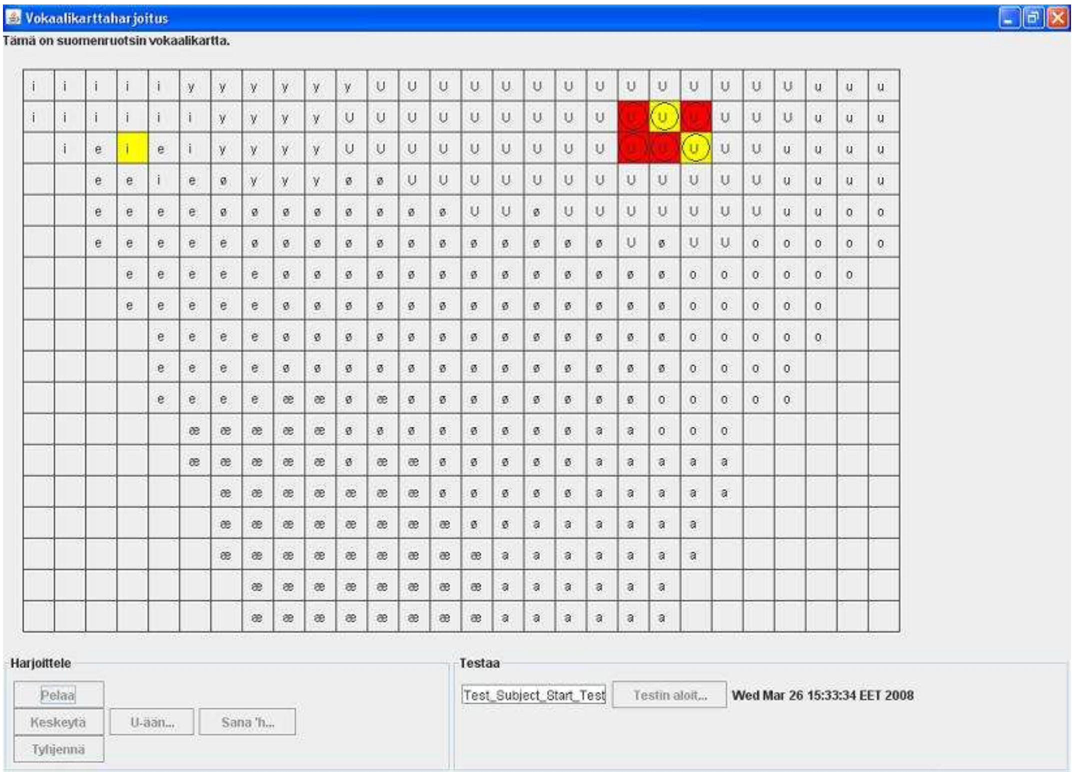


Figure 9 Hits in The Vowel Game's vowel diagram

4.2 Accent Lab

Accent Lab was one of the vowel training systems, which was designed to utilise different types of visual methods to help evaluate the student's performance. Among other things, it measures and displays a diagram illustrating the pitch and the volume of the student's voice and then compares it to a model programmed by the teacher or the pronunciation coach. By comparing the two diagrams, the student can see how well he/she succeeded in replicating the correct emphasis/stress and pitch.

Accent Lab also evaluates the student's vowels by utilising spectrograms, but unlike The Vowel Game, it does not provide real-time formant tracking. The formants of a vowel are displayed both in the spectrogram and in a vowel diagram. As explained earlier, formants are energy concentrations in a frequency spectrum. With vowels, only the first two formants, F1 and F2, are relevant with respect to the spectrogram and the vowel diagram. The vowel diagram can be used to illustrate the location of the target vowel and how far or close the student's vowel falls to that target. In this respect, a graphic real-time interactive vowel diagram is tangible in the sense that it can be modified/programmed to clearly display the distance between the target vowel and the student's production of the vowel in question. A spectrogram, where the actual formants are displayed as darker areas, may not be as informative as the vowel diagram and would certainly require spectrogram interpretation skills (Carey 2002: 163). The use of this system therefore requires a tutor in case the student is a not an expert in phonetics and spectrograms (Healey 1998). Accent lab is available at www.accentlab.com (Accent Technologies 1997).

4.3 KAY Sona-Match

Another similar product is KAY Sona-Match. It resembles Accent Lab with the exception that it performs the formant analysis automatically. It utilises the LPC algorithm (also explained in more detail Chapter 3) to determine the first two formants and displays the vowel in a vowel diagram. Kay Sona-Match also provides feedback in real-time like The Vowel Game. The only inadequacy of the system is that it does not teach the student how to improve his/her performance. This is due to the fact that it does not give any specific instructions on how to adjust the first two formants by changing the position/shape of the tongue or the lips. KAY Sona-Match is available at www.kayelemetrics.com (KayPENTAX 2008).

4.4. The Video Voice

Another more recent product is The Video Voice available from www.videovoice.com (Micro Video Corporation 2008), which has several visual representations of both the student's productions and the various models. It has different formant displays for vowels and P-A-R (Pitch-Amplitude-Rhythm) displays for the different 'invisible aspects' of speech such as rise-fall or fall-rise patters and how loudly or softly sounds are produced. The Video Voice also includes pitch and intensity measurement methods and games for speech practise aimed at younger students or children in need of speech therapy. The problem of the Video Voice is that, unlike The Vowel Game, it does not provide feedback in real-time.

4.5 The 3-D Vocal Tract Project

All the other examples in this chapter, which utilise a vowel diagram, are based upon a 2-dimensional representation of the vocal tract. However, 3-dimensional models have been developed as well. For example, www.speech.kth.se/multimodal/vocaltract.html is an application developed by KTH (Kungliga Tekniska Högskolan) and it offers a 3-D model of both the vocal tract and the nasal track along with the tongue, the lips and the teeth. All these parts are represented with colours that resemble the actual colours of these speech organs. The main difference between this system and The Vowel Game is, as stated earlier, the 3-dimensional aspect and the fact that the 3-D project does not display a vowel diagram as such. The application is presented on a website (Kungliga Tekniska Högskolan 2005).

4.6 Trollerilådan

Trollerilådan is a Swedish speech training system that was developed within the EU project “SPECO” between 1998 and 2001 for children between 4-10 years with speech impediments and hearing-impairments (Öster 2006: 167). However, it can also be used for adults. The difference between The Vowel Game and Trollerilådan is that the use of the latter is not solely limited to the training of vowels. Trollerilådan (‘Box of tricks’) has been divided into four segments: sound preparation, vowels, sibilants and intonation where the sound preparation segment is developed for the training of loudness, spectrum, pitch and rhythm (Eriksson 2004: 15). Trollerilådan is available from Frölunda data (Frölunda data 2007).

4.7 Pronunciation power

The main difference between Pronunciation Power and The Vowel Game is that Pronunciation Power does not feature a vowel diagram, but displays the student's sound (e.g. a vowel) in the wave audio format instead. This application also offers the option of comparing one's sound to the sound of the instructor. Carey deems to use of the wave audio format rather ineffective since it only displays the intensity of the speech signal. With this format, variables of speech rate, articulatory settings and airflow generate a different result each time even in the case where the original model speaker attempts to recreate the model (Carey 2002: 160). Healey also concludes that the waveforms may be difficult to interpret by the students on their own (1998). Pronunciation power is available at www.englishlearning.com (English Computerized Learning, Inc. 2006).

4.8 Ellis Master Pronunciation

The Ellis Master Pronunciation displays the instructor's vocal tract and the movements of the articulators and compares them to the corresponding movements of the student. This application also offers the option of listening to a male or a female voice and then observing the possible differences in the movements of the articulators. The difference here between the application and The Vowel Game is that the system does not display the vowel diagram as such, but offers a side view of the actual vocal tract instead. The weakness of this system is that the student may find it difficult to understand how he/she should alter the position of his/her articulators (e.g. tongue) to improve the pronunciation since the student is unable to see inside his/her vocal tract (Carey 2002: 160). Also, the record-keeping facilities of this system require that each student has his/her own ID provided by the teacher (Healey 1998). The Ellis Master Pronunciation is available at www.formavision.com (CALI, Inc & Formavision 2000).

4.9 Protrain Vowel Space Display

The Protrain Vowel Space Display features a vowel diagram and also displays the formant values. It also offers the option of recording the student’s pronunciations for later analysis. The weakness of this application is that, unlike The Vowel Game, it does not provide feedback in real-time and that the formant axes have not been inverted to correspond with the tongue movements. Also, it does not display true IPA fonts (Carey 2002: 164). Protrain is available at www.protrainsys.com (ProTrain 2005).

4.10 Dr. Speech

Dr. Speech is slightly similar to Protrain in the sense that it displays a vowel diagram. In addition to the diagram, it also displays the sound in the wave audio format. Dr. Speech is available in two versions. They both provide real-time feedback of F1-F2, but the Speech Training version also provides the option of viewing two speech samples at the same time in the wave audio format (Carey 2002: 165). Both are available at www.drspeech.com (Tiger DRS, Inc. 1999).

4.11 Technology Enhanced Accent Modification (TEAM)

The Technology Enhanced Accent Modification (TEAM) displays not only the student’s hits in a vowel diagram, but also e.g. an audio model in the wave audio format, which the student may try to match through imitation. TEAM is essentially a reconfigured version of Dr. Speech and the problem with it is that the vowel model templates cannot be altered (Carey 2002: 161). The similarity between the system and The Vowel Game is simply the use of the vowel diagram. The Technology Enhanced Accent Modification is available at www.fipse.aed.org (Fipse grant database 2005).

4.12 Speech Viewer III

The original Speech Viewer was developed by IBM in 1983 and its target areas for speech training are voicing, F0, speech timing, loudness and vowel production (Eriksson 2004:16). The application is clearly directed towards children, but the authoring of this system is non-intuitive unless one is an expert or a speech clinician as assessed by D. Healey (1998). Speech Viewer III consists of 13 different programs and A-M. Öster evaluated the vowel diagram program to be a useful pedagogical aid in displaying the students' vowel deviations. She also deemed the vowel diagram to be an efficient tool in describing the tongue's role in vowel production (2006: 160). This finding slightly contradicts M. Carey's view on vowel diagrams as in the case of Kay Sona-Match. Speech Viewer III is available at www.ibm.com (IBM 2008).

4.13 Praat

Praat is a freeware phonetics program that can be downloaded from www.fon.hum.uva.nl/praat (Boersma and Weenink 2008). Praat is not a vowel training system as such, but it can be used to record vowels and to display the vowel's spectrogram illustrating the formants as shown in Figure 10. The problem with Praat lies in the fact that it is a complex system directed towards specialists who are skilled in the areas of the LPC algorithm and formant and spectrogram analysis. In this sense, it is certainly not easy for a beginner to use and for someone who is not knowledgeable about the principles of formants and spectrograms. A spectrogram and the wave audio view is illustrated in Figure 10 and the commands are issued in the interface depicted in Figure 11.

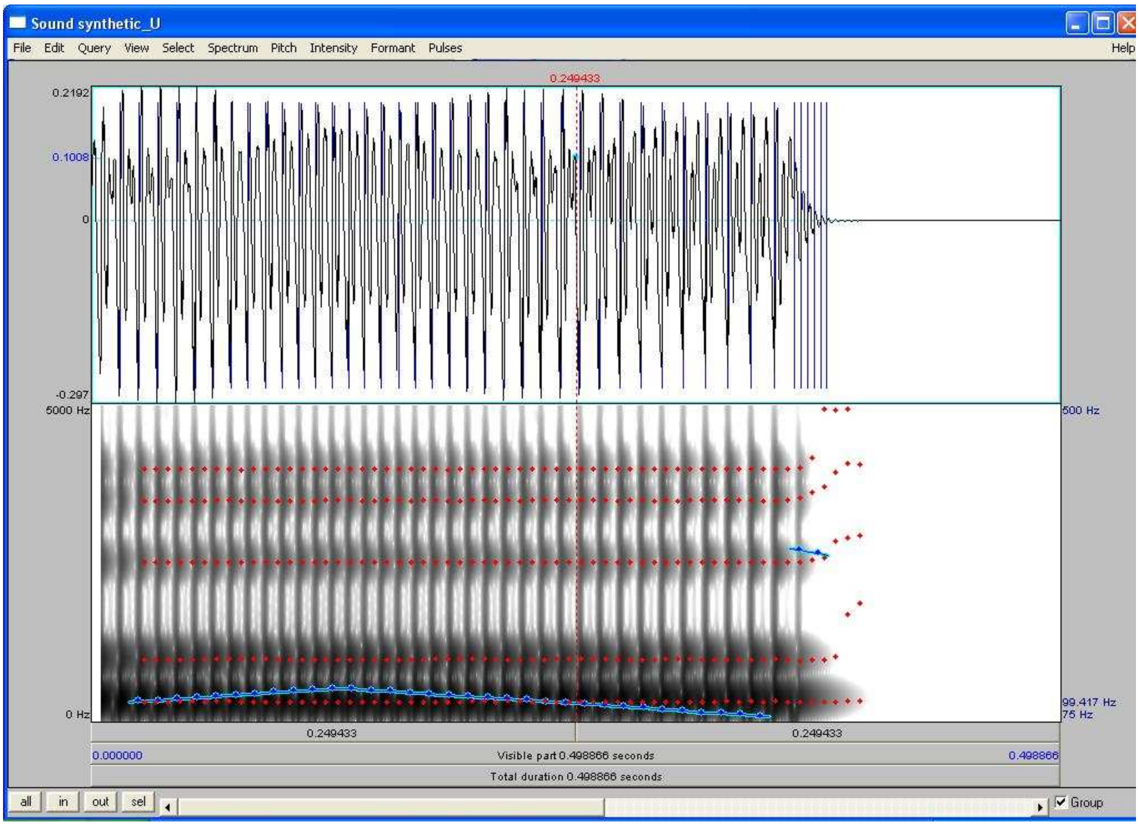


Figure 10 Praat, spectrogram and wave audio view (Boersma and Weenink 2008)

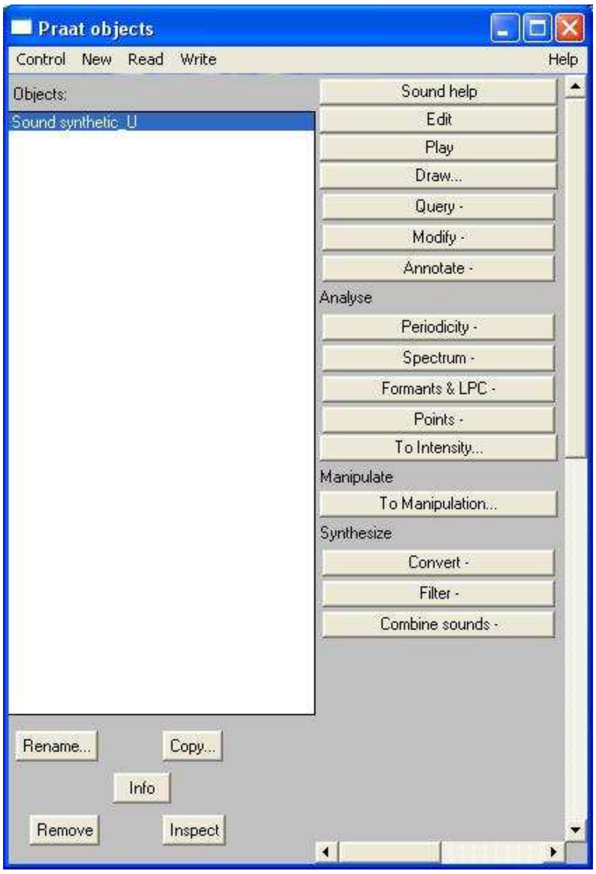


Figure 11 Praat's interface (Boersma and Weenink 2008)

4.14 A Summary of the systems and their characteristics

What appears to be a common denominator for most of the systems presented in this chapter is that they often offer an F1/F2-plane based on visual feedback. The planes display the vowel hits either in real-time as in the case of The Vowel Game or with a delay. While the use of such vowel diagrams can be informative as stated by M. Carey, they often fail to provide the student with specific information on how the vowel should be pronounced and how the vocal tract and its articulators should be configured. For instance, they do not explain explicitly how the position of the tongue should be altered or when the lips should be rounded. Other systems, such as Ellis Master Pronunciation, offer a side view of the vocal tract. Again, this setting presents yet another problem, since the students are unable to see within their own vocal tracts and to observe how the positions of their articulators change. This is also the case with The Vowel Game since it does not offer any information on how the lips should shaped. As M. Carey suggests, such systems often require the presence of a speech clinician, a therapist or a phonetics specialist to aid the student in the learning of new vowel categories. At least, the presence of such an expert is deemed vital when the student starts to use the system for the first time.

Another problem with some of the systems seems to be their complexity and their limitations with regard to authoring like in the case of Ellis Master Pronunciation. Systems such as Accent Lab, Praat and even Speech Viewer III, which is a speech therapy software developed for children, require advanced knowledge of speech production. Praat, in particular, requires spectrogram analysis skills and it is therefore not a suitable tool for a beginner. Some of the systems also fail to provide the IPA fonts, which is the case with Protrain.

However, most of the systems offer various others methods of speech training or therapy besides vowel training such as pitch and intensity measurement methods or intonation pattern exercises. In this sense, some of the systems such as The Video Voice, Trollerilådan or TEAM can be deemed more elaborate than The Vowel Game whilst keeping in mind that The Vowel Game is still under development.

The next chapter centers on the testing of The Vowel Game and describes the basic test settings and the test methods. Chapter 6 provides the information on how the data was composed of and how it was treated before the analysis. Chapter 7 provides the various analysis methods applied to the data and finally Chapter 8 provides the conclusions and evaluations based on the analyses.

5. The Testing of The Vowel Game

The purpose of the testing of The Vowel Game is to show the effect of the real-time visual feedback of the game. This type of an effect, based on real-time visual feedback, has been studied earlier by Dowd, Smith and Wolfe using another type of acoustic measurement of the vocal tract (1998: 1). In the testing of The Vowel Game, the subjects were advised to imitate a vowel according to audio models and then consequently produce the vowel in The Vowel Game. With this method, the subjects received immediate feedback on how far or close their vowels hit the target/reference point. With successive repetitions of the vowel within one test and over many consecutive exercises, the standard deviations of their vowels in the diagram could be perceived along with their average concentrations in the vowel diagram.

The tests involved two groups; the treatment group and the control group. Both groups practised the imitation of the sustained synthetic /ʉ/-vowel and the word /hʉs/ according to audio models. The general difference between the groups was that only the treatment group was able to utilise The Vowel Game's visual real-time feedback in the imitation process. These types of experiments, involving the imitation of vowels e.g. in a word, have been conducted as early as 1952 by G.E. Peterson and H.L. Barney (1952: 175). Essentially, they studied the variation of vowels in speech production using over 70 speakers and listeners and several 3-letter words including a vowel. The general difference between the tests of Petersen and Barney and the tests of this study was that the Petersen and Barney test subjects first recorded their own vowel productions and then consequently imitated these productions whereas the subjects in this study imitated pre-recorded models. The Petersen/Barney tests also involved the identification of the vowels whereas this study involved only the production/imitation of the sustained vowel /ʉ/ and the word /hʉs/. No identifications or discriminations were involved. Another example of a study involving the imitation of vowels has been the study of Dowd, Smith and Wolfe (1998: 1).

As for the questions, whether or not imitation can aid pronunciation and what is the cause for the learning with imitation, one of the plausible explanations for the learning process may be, as explained in the introduction, the mirror-neuron system in humans.

This theory has been suggested by Rizzolatti and Craighero (2004: 169) and Arbib (2002: 272). Essentially, the function of the mirror-neurons can be explained in the following manner. The mirror-neurons are located in the Broca area on the brain and a certain activity perceived by the individual, e.g. pronunciation or speech, stimulates the mirror-neurons, which, in consequence, activates the neurons and aids the person in the imitation of the perceived action. However, it should be pointed out that the purpose of this study was not to prove the existence of the mirror-neurons or to show that they caused the learning in imitation, but merely to suggest that the mirror-neurons might be one plausible explanation for the learning process.

Also, what should be stressed is the fact that the purpose of the tests was not to teach the subjects native-like pronunciation but rather to investigate if The Vowel Game's real-time visual feedback has an effect in making the pronunciation more constant. In any case, researchers have suggested that native-like pronunciation can be learned approximately between the age of 2 and the early teens with sufficient exposure to the language (see Chapter 2.6). In the testing of The Vowel Game, all the test subjects were adults.

In any case, The Vowel Game's vowel diagram has been coded as a constant area and therefore all the subjects cannot be assumed to hit the /ʉ/ category's center. This is due to the fact that the shape, size and the position of the vowel diagram vary according to the speaker. For instance, the shapes, sizes and the positions of the vowel categories of two different speakers are rarely, if ever, the same, especially when the sex is taken into account. Each and every speaker has his/her individual vowel diagram. Rosner and Pickering state that the productions of one vowel by different speakers can be very diverse spectrally due to the differences in the vocal tract size and shape (1994: 265). It is therefore beneficial to study how the deviation and the area of the subjects' plotted vowels change over time due to the effect of the real-time visual feedback. Do they concentrate near the vicinity of the target vowel's category or do they scatter over time? Also, what happens to the vowels of the control group when they use The Vowel Game without ever seeing the graphic vowel diagram and without benefiting from the real-time visual feedback? Do the plotted vowels of the control group scatter more than the ones of the treatment group that has practiced the pronunciation of the /ʉ/ vowel with

The Vowel Game 4 times over a period of 1 week? Also, what are the changes in the average F1, F2 and the Euclidean values within and between the groups? Is there a perceivable difference in the average values of the treatment group between the start test and the end test and how does this change compare to the performance of the control group?

5.1 The Test method for the treatment group

As stated earlier, the testing of The Vowel Game involves two test groups. The first is the treatment group, which consists of 10 Finnish-speaking adults (6 males and 4 females) that have received generic language teaching in Swedish at school. The other group of 10 Finnish-speaking adults (6 males and 4 females) functions as the control group. The reason for selecting adults for the tests instead of children is related to the problem of the children's higher formant values as explained in Chapter 2.1. Both the treatment group and the control group use The Vowel Game on 4 separate occasions, approximately 3-4 days apart. The difference between these two groups is that the control group never sees The Vowel Game although their vowel exercises are recorded with The Vowel Game and they are advised to repeat the same procedures as the treatment group (see Figure 12). The approach in the study is similar in principle to the approach Dowd, Smith and Wolfe took in their study. They used a visual display of the impedance spectrum of the vocal tract as real-time feedback. This way the test subjects could understand the vocal tract configuration required to pronounce the target vowel (1998: 1). Essentially Dowd, Smith and Wolfe also used two groups, one of which functioned as the control group. The concept of using a treatment group and a control group was also utilised by M. Carey and A-M. Öster in their respective studies.

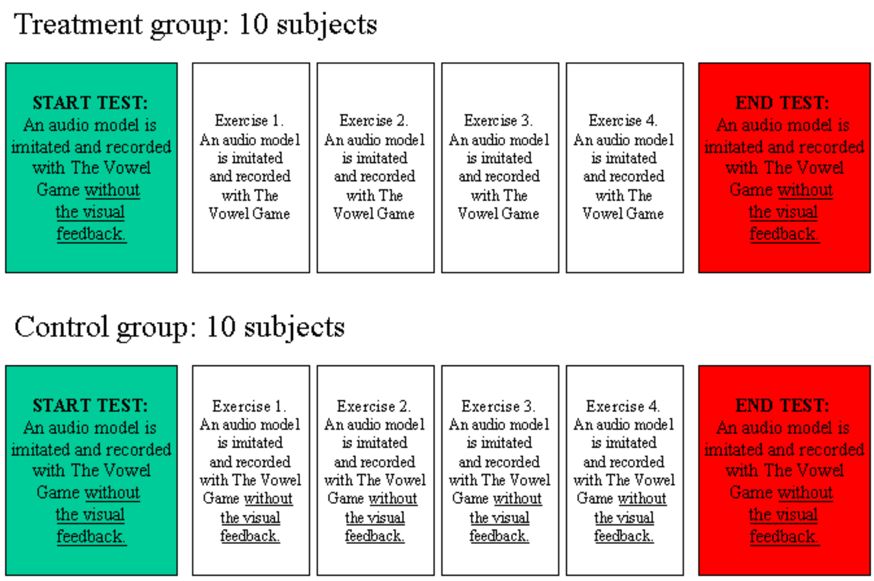


Figure 12 The test structure

The first segment of the testing is the start test in which the level of the two groups could be observed. The start test utilised a test recording, which consisted of 10 synthetic models of the vowel /ʉ/ and 10 natural models of the Swedish word ‘hus’. This method of using both natural speech and synthetic material is the same used by e.g. Taisto Määttä in his contrastive studies in the perception of the vowel sounds of Swedish by Finnish Speakers (1983: 187). In the test recording used for The Vowel Game, there was a 3 second gap between all the models so that the subject had a sufficient amount of time to react and imitate the model. Both groups were advised to listen to the test recording through a set of headphones (KOSS 45SB) and immediately imitate the sound when it was played back. In other words, repeat after the tape. The start test was short and only lasted approximately 80 seconds. The altogether 20 imitations (10 x /ʉ/ and 10 x /hʉs/) were recorded in text format with The Vowel Game and stored in a text file. The Vowel Game’s vowel diagram was not visible to the subjects at this stage. This start test was identical to both groups and its purpose was to indicate and show how the subjects perform in The Vowel Game’s vowel diagram before the exercises and the end test. An alternative way of determining the levels of the two groups before the start test and the end test could have been a procedure, which is described below. The method is very similar to the one suggested by A-M. Öster, which involves an evaluation by a speech therapist (2006: 197). The method was abandoned due to the fact that the fundamental idea of the tests was not to evaluate the skills of the students, but to test whether or not The Vowel Game can guide and aid pronunciation.

The omitted procedure:

> Before the actual testing commences, the pronunciation skills of the participants are tested using a piece of text in Swedish, which the participants are advised to read out loud. The performances are videotaped so they can be analysed later by a phonetics specialist. The performances are evaluated using a scale of 1-5 (1=poor, 5=excellent). This is carried out to determine the basic level of the group. If a certain treatment or control group member(s) is deemed to be below the average level of the group, the member is replaced with a more suitable test person. This is to ensure that the skills of the group are as homogenic as possible to avoid any drastic deviations in the actual test results. The same evaluation test is carried for the control group.

After the initial start test described earlier, it was time to perform the exercise sessions. The procedures in the sessions were the following. The testing took place in a specifically designed usability laboratory. The laboratory room was divided into two areas: the testing side and the observatory. These two areas were separated by a sound-proof glass, which enabled the supervisor of the test to observe and monitor the setting. The idea of this setting was that the presence of the supervisor should not disturb the test subject. The test subject was seated in front of a computer screen and given a pair of headphones with a microphone (KOSS SB45). He/she was briefly instructed on how to operate The Vowel Game; how to playback the audio model and how to commence the recording of the test. The audio level for the models was kept constant for all the subjects. The test person was then left alone. The entire test procedure was videotaped for further analyses and evaluations.

Figure 13 The test setting: The Vowel Game, headphones + microphone, video camera, noiseless environment (the window separating the test side and the observatory is sound-proof)

Figure 14 The observatory and the video mixing facilities

1) The test subject is advised to listen to an audio model of the target vowel /ʉ/ and then to imitate the sound to the best of his/her ability. The played audio model can be listened to by pressing a 'play'-button in The Vowel Game's interface. The subject has a few minutes to try to imitate the vowel until the recording commences. The subject is also advised to listen to an audio model of the word 'hus' (/hʉs/) and to imitate the word. This is, in essence, a pre-training phase.
2) Before the actual recording is started, the supervisor enters the info regarding the test subject's name and the test session into the text field. This helps to identify the recording results later when they are examined in the text file.
3) When the test subject is ready and the supervisor has given an 'ok'-sign through the window, the subject presses the record-button in The Vowel Game, which commences the same test recording that was used in the start test. The subject imitates the 10 /ʉ/-vowels and the 10 /hʉs/-words of the test recording, which are consequently recorded with The Vowel Game. The test is finished when the 10^{th} and last 'hus'- word has been imitated.
4) After the session, the supervisor provides the subject feedback on how he/she could improve the pronunciation by changing the position of the lips or the tongue.

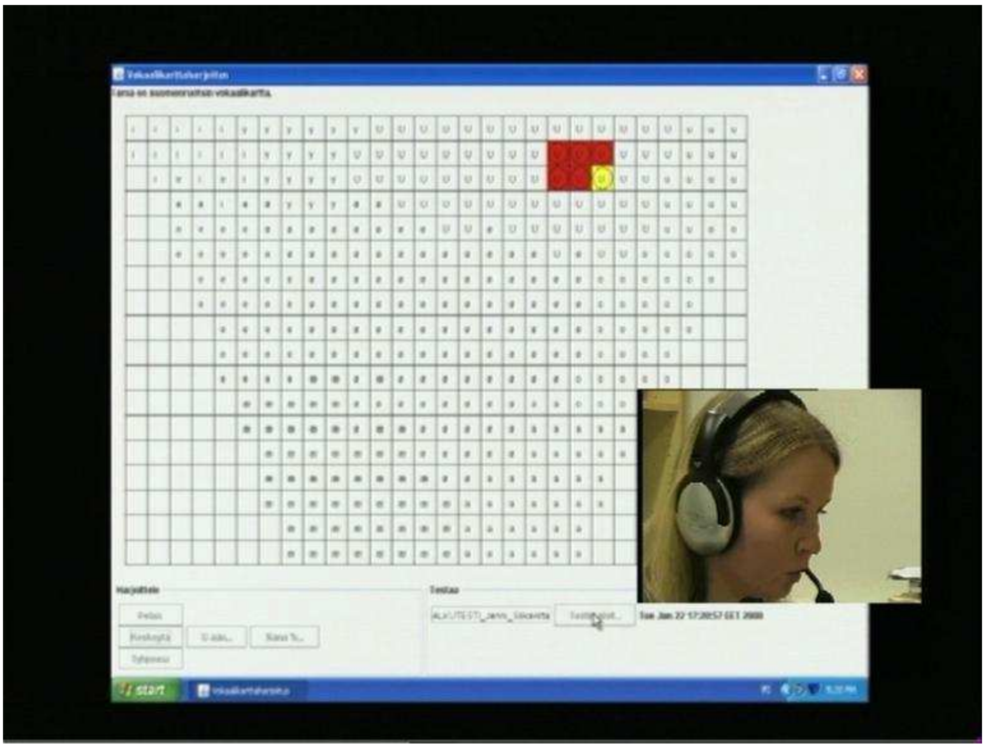


Figure 15 A treatment group test subject pronouncing the /ʉ/-vowel in the start test

Figure 15 illustrates a test subject of the treatment group in the start test. At this point The Vowel Game's vowel diagram was obscured from the test subject by a piece of cardboard, but in the video mix both The Vowel Game and the subject are visible. The pronounced /ʉ/-vowel can be seen in the diagram as a yellow box on the right side of the reference box.

The overall duration of one exercise session is approximately 5-10 minutes and the practice session is repeated 4 times over a period of approximately 1 week. The first of the 4 exercises is carried out immediately after the start test to save time and the last is followed immediately by the end test, which is identical to the start test. The method of keeping all the test sessions and the exercises short (max. 15 minutes) coincides with the criterion set by A-M. Öster, which defines that a training session should not exceed 30 minutes in order to be efficient (2006: 197). According to her theory, a complete training-period should include approximately 10 sessions before the final evaluation. However, within the constrictive schedule of The Vowel Game's testing, such a number of sessions was deemed impossible. The functionality of the entire test setting was first verified in a pilot test using a single test person to ensure that the actual tests could be performed without any technical problems.

In the statistical analyses, the start test and the end test results are compared between the groups. What is particularly interesting is to compare the end test results of the treatment group and the control group.

5.2 The Test method for the control group

The procedures for the control group are identical to the procedures of the treatment group with one crucial exception. The control group is only advised to listen to the audio model of the vowel or the word and to imitate it to the best of their ability. The control group never sees the actual Vowel Game and is therefore never exposed to the effect of the real-time visual feedback. This is to ensure that the possible effects of The Vowel Game's real-time feedback to the treatment group could be perceived as a measurable difference between the two groups. The control group imitates the vowels/words according to the audio model without the visual feedback of the application. However, their imitations of the vowel /ʉ/ are nevertheless recorded with The Vowel Game using the same test recording that is used in all the tests.

Figure 16 The control setting (Notice the cardboard obscuring The Vowel Game's vowel diagram)

5.3 The Data parameters and the hypothesis

When the record button has been pressed, The Vowel Game checks every 32 ms if there were hits in the vowel diagram. In the case of a hit, The Vowel Game records it in a specified text file.

The game has been modified for the sake of the test to record the Euclidean distance from the center of the target area. The center point of the target area is the following:

F1 = 300 Hz

F2 = 969 Hz

The Euclidean distance is of the form:

$$d = \sqrt{\left(F1 - F_{1,test}\right)^2 + \left(F2 - F_{2,test}\right)^2} \quad (46)$$

$F_{1,test}$ in Equation 46 is the first formant of the test person's pronounced vowel and $F_{2,test}$ is the second.

Equation 46 is fundamentally the same as the one used by Boe, Schwartz and Vallee (1994: 195) in their study with one exception. Instead of using F2, they used F'2, which is a computation of F2, F3 and F4. The distance gives information on how far/close the subject's vowel hit the center of the target area. Since the diagram is not calibrated according to the subject's voice and vocal tract, it is impossible to assume that the optimal target vowel of each subject would be at the center of the target area. As Rosner and Pickering stated, the productions of one vowel by different speakers can be very diverse spectrally (1994: 265). Still, the distance gives valuable information of the relative position of the subject's vowel to the center of the target area. What is interesting to observe is the mean value of the distance of the ten hits, their standard deviation and the coefficient of variation. Of course, it is important to note that the distance alone does not provide any definitive information of the exact locations of the hits. The Euclidean distance only provides the radius from the center and the hits could be located on any single point on the circle defined by the radius (which is the distance). Still, the concentration of the hits is always verified by examining the average hits in each diagram to ensure that the hits are not scattered. Also, if the average value for the Euclidean distance has decreased from the start test, we know at least that the average hits have drawn closer to the reference point. In the case of the treatment group, this would also indicate that the real-time visual feedback has aided the test subject to direct his/her hits closer to the target.

In any case, the mean value is of the form:

$$E = \frac{\sum_{i=1}^{n} x_i}{n} \tag{47}$$

The distance d in Equation 47 has been notated as x and the number of the hits as n.

The standard deviation is of the form:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \tag{48}$$

where the mean value E has been notated as $\bar{x}$.

Finally, the coefficient of variation is of the form:

$$CoV = \frac{s}{E} = \frac{\frac{\sum_{i=1}^{n} x_i}{n}}{\sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}} \tag{49}$$

The student's hits are recorded in a text file (.txt) in the following format:

Date and time; a text field; F1; F2; the vowel; the Euclidean distance.

E.g:

Thu Sep 13 19:20:49 EEST 2007; text_field;656;1406;ø;563

In the above example, the subject's vowel parameter values are:

F1 = 656 Hz, F2 = 1406 Hz.

The vowel is /ø/ and the Euclidean distance according to Equation 46 is 563 Hz.

When the subject imitates the vowel /ʉ/ in a single test, rows indicating the hits are recorded in the text file in the following manner:

Thu Sep 13 19:20:49 EEST 2007;text_field;656;1406;ø;563
Thu Sep 13 19:20:49 EEST 2007;text_field;656;1406;ø;563
Thu Sep 13 19:20:49 EEST 2007;text_field;656;1406;ø;563
Thu Sep 13 19:20:49 EEST 2007;text_field;656;1406;ø;563
Thu Sep 13 19:20:49 EEST 2007;text_field;656;1406;ø;563

The mean value, the standard deviation and the CoV can be calculated for F1, F2 and the Euclidean distance according to Equations 47, 48 and 49. When the subject repeats the same test four times as depicted in Figure 12, we receive four sets of mean values, standard deviations and the CoVs. Within the scope of this study, only the start test and the end test results will be examined and compared with only some attention to the performance in the 4^{th} exercise, which is executed immediately before the end test.

What is intriguing to observe is how the CoVs of the hits change between the start test and the end test due to the proposed effect of the four exercises between them. Have the CoV-values decreased in the end when compared to the start test? The hypothesis is that the CoV-values of the F1-and F2-values of the treatment group will be smaller than the corresponding CoVs of the control group. This is due to the assumption that the immediate real-time visual feedback of The Vowel Game aids the subjects in the treatment group in producing vowels, which hit the same area in the vowel diagram more constantly than the control group, which does not receive any real-time visual feedback. Here it is crucial to illustrate to the test subjects that they should always attempt to imitate the audio model to the best of their ability regardless of whether or not they hit the target area of /ʉ/ dead center. This is due to the fact that the diagram is not calibrated and that the target area's center is not an absolute model. Every test subject will have their unique optimal 'spot' in the diagram, which best represents their /ʉ/. The position of this /ʉ/ is not necessarily the center of the target area.

After the start test, each test subject in the group will have their average, standard deviation and the coefficient of variation (CoV) values for the F1, F2 and the Euclidean distance for each of the 10 /ʉ/-imitations and 10 /hʉs/-imitations. The /ʉ/-and /hʉs/-sections will be analysed separately. This means that with one group, there will be 100 values (10 subjects x 10 imitations) in the /ʉ/-section for each parameter (F1, F2 and the Euclidean distance). There will also be 100 corresponding values in the /hʉs/-section as well.

It is also possible to calculate an average, a standard deviation and a CoV for a single test subject for the 10 imitations of e.g. /ʉ/ in the start test and calculate the corresponding values in the end test as well. By comparing these two sets of averages, it is possible to evaluate a single test subject's performance between the start test and the end. The comparison must be made for both the /ʉ/-section and the /hʉs/-section.

The four exercises are performed after the start test. When they have been executed successfully, it is time to perform the end test, which is identical to the start test. In the end test, neither the treatment group nor the control group see The Vowel Game's vowel diagram, but their imitations of the /ʉ/ and /hʉs/ are nevertheless recorded with The Vowel Game. The end test will be evaluated on its own and separately from the four exercises. Here the hypothesis is that the four prior exercises will cause a measurable difference between the performances of the treatment group and the control group.

In the end test, the CoVs of the F1 and F2 values will be calculated as described earlier, based on the results from the end test. The assumption is that in the end test, the CoVs of the F1, F2 and the Euclidean distance values of the control group will be larger than those of the treatment group. All this is examined in detail in Chapter 7.

The data from the start test and the end test from both the treatment group and the control group will be analysed using a statistics software SPSS. The first test to be performed is the Kolmogorov-Smirnov-test to see if the data has normal distribution. This test defines whether the data should be analysed using ANOVA (Analysis of variance) or a non-parametric equivalent of the t-test, A Mann-Whitney U-test. If the data does not have normal distribution, the ANOVA analysis must be abandoned.

6. The Results and their treatment

Before any analysis could be performed, the data of the test subjects was examined and all the false hits (outliers) had to be removed from the data to prevent them from affecting the actual results. There were essentially four fundamental factors, which contributed to the false hits.

1) The first problem is related to test subjects that have a high F0. The weakness of the LPC algorithm and the all-pole-model is the energy dependence on the maxima in the spectral envelope and the exceeding difficulty of predicting the actual formants correctly instead of predicting the strong multiples of F0 (harmonics) (Alku and Bäckström 2004: 95). In many cases, the female voices produced false hits in the /i/, /e/ and /ae/ regions due to the fact that the multiples of their F0 were stronger than their actual formants. This was apparent especially in the case of the F2 values when the value was predicted incorrectly due to a strong multiple of F0 (harmonic) slightly above 2000 Hz.

2) The second problem is the phenomenon known as merging. The first two maxima in the spectral envelope are so close to one another that The Vowel Game perceives them as one and again selects the next visible maximum, which is F3, and interprets it as F2. Again this leads to a false hit (Paganus et al. 2006: 700).

3) The third problem is related to the occasionally low amplitude of F2. The amplitude of the second maximum in the spectral envelope may occasionally be so low that the program does not interpret is as a maximum and instead selects the third maximum as the second (Paganus 2007: 58).

4) The fourth problem is related to noise. The microphone is able to detect the slightest interferences in the room caused by ventilation, processor noise or other sources and consecutively plots the interferences in the diagram as vowel hits. This occurs even in the case, when the Microphone level in the Windows XP system has been set at mid level. This is a phenomenon, which requires further development and attention in the future versions of The Vowel Game.

Table 2 The Vowel Game's output in text form

Month:Day:Time:					Year:	Text:	F1(Hz):	F2(Hz):		Dist:(Hz)
Mon	Jan	21	10:40:17	EET	2008	START_TEST_Test_subject_8	0	0	i	1014
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	312	1031	U	63
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	312	2250	i	1281
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	312	812	U	157
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	312	937	U	34
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	312	906	U	64
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	312	906	U	64
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	281	937	U	37
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	281	937	U	37
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	281	906	U	65
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	281	906	U	65
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	281	906	U	65
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	250	875	U	106
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	250	781	U	194
Mon	Jan	21	10:40:20	EET	2008	START_TEST_Test_subject_8	0	0	i	1014

Test subject 8. pronounced /ʉ/ at 10:40:18 on the 21st January 2008 in the start test. The pronunciation at 10:49:18 has produced 13 samples/hits in text file meaning that the vowel's duration was 13 * 32ms = 416ms. The first sample appears on the second row in Table 2. On the third row appears the second sample, where the F2 value has suddenly increased from the average value of 1006 Hz to 2250 Hz. This is a phenomenon where The Vowel Game has either interpreted the third peak in the spectral envelope as F2 or where the LPC algorithm has favoured a multiple of F0 at 2250 Hz over the correct F2. Therefore, we have received a batch of hits, where the rest are in the /U/-region and one is in the /i/-region. To prevent this one false hit from having an effect on the standard deviation of the F2 values, the false hit was removed. Also, the Euclidean distance of the false hit was 1281 Hz from the center of the target area. In the test data, all the hits that had an Euclidean distance of over 1000 Hz were eliminated from the results. This was to prevent the 4 listed problems from having an effect on the results. The false hits or outliers were located primarily in the /i/, /e/ or /ae/-categories, due to the fact that the F2 peak had been selected incorrectly from the spectral envelope.

The false hit on the third row has been removed from the data as illustrated in Table 3.

Table 3 The exclusion of a false hit

Month:Day:Time: Year: Text: F1(Hz):F2(Hz): Dist:(Hz)

Mon	Jan	21	10:40:17	EET	2008	START_TEST_Test_subject_8	0	0	i	1014
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	312	1031	U	63
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8				
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	312	812	U	157
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	312	937	U	34
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	312	906	U	64
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	312	906	U	64
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	281	937	U	37
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	281	937	U	37
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	281	906	U	65
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	281	906	U	65
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	281	906	U	65
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	250	875	U	106
Mon	Jan	21	10:40:18	EET	2008	START_TEST_Test_subject_8	250	781	U	194
Mon	Jan	21	10:40:20	EET	2008	START_TEST_Test_subject_8	0	0	i	1014

In some cases there were more than one false hit in a batch of hits, which had to be removed. This was probably an effect of the merging of the first to maxima in the spectral envelope. In such a case, The Vowel Game falsely interprets the third peak as the second. In unclear cases, the video mix of The Vowel Game’s vowel diagram and the test subject was examined to verify that all the hits in a batch were authentic hits and not caused by noise or other interferences. In some instances a hit had to excluded from the results altogether. This was due to the first problem related to the weakness of the LPC algorithm and its tendency to favour high-energy regions of the signal spectrum (e.g. the multiples of a high F0) as stated by Paavo Alku and Tom Bäckström (2004: 95). An example of this phenomenon appears in Table 4.

Table 4 The exclusion of an entire hit

Month:Day:Time: Year: Text: F1(Hz):F2(Hz): Dist:(Hz)

Mon	Jan	7	13:03:33	EET	2008	START_TEST_Test_Subject.7	0	0	i	1014
Mon	Jan	7	13:03:33	EET	2008	START_TEST_Test_Subject.7	0	0	i	1014
Mon	Jan	7	13:03:33	EET	2008	START_TEST_Test_Subject.7	343	2562	e	1593
Mon	Jan	7	13:03:33	EET	2008	START_TEST_Test_Subject.7	406	2562	1596	
Mon	Jan	7	13:03:33	EET	2008	START_TEST_Test_Subject.7	406	2593	1627	
Mon	Jan	7	13:03:33	EET	2008	START_TEST_Test_Subject.7	375	2250	e	1283
Mon	Jan	7	13:03:33	EET	2008	START_TEST_Test_Subject.7	375	2531	e	1563
Mon	Jan	7	13:03:33	EET	2008	START_TEST_Test_Subject.7	375	2375	e	1407
Mon	Jan	7	13:03:33	EET	2008	START_TEST_Test_Subject.7	281	2156	i	1187
Mon	Jan	7	13:03:33	EET	2008	START_TEST_Test_Subject.7	375	2468	e	1500

In Table 4 test subject 7. imitated the sustained /ʉ/-vowel in the start test on the 7th January 2008 at 13:03:33. When the video of this particular imitation was examined, it was clear that the female test subject had pronounced /ʉ/. However, when the hit is examined in Table 4, all the hits are at a distance of over 1000 Hz from the reference point and they appear to be in the /i/ and /e/-regions. Also, all the F2-values are well over 2000 Hz. This is an indication that the LPC algorithm has favoured a multiple of the female test subject's F0 instead of the correct F2. Alternatively, F1 and F2 have emerged in the spectral envelope and the program has interpreted F3 as F2. In either case, this hit and all other hits alike had to be excluded from the analyses.

After removing the false hits from the batches, the average values for F1, F2 and the Euclidean distance from the center of the target area were calculated for each batch of hits. Therefore, when a subject imitated the /ʉ/-vowel 10 times, he/she produced 10 batches of hits. The average values for F1, F2 and the Euclidean distance were calculated for each of these batches, thus forming a group of 10 averages. In the example in Table 3, the average values for F1, F2 and the Euclidean distance were calculated as shown in Table 5 according to Equations 47, 48 and 49.

Table 5 The average F1, F2 and the Euclidean distance values of a sustained vowel

F1(Hz):F2:Hz) Eucl.Dist.(Hz)

0	0	1014	i		F1 (Hz)	F2 (Hz)	Eucl.Dist.(Hz)
312	1031	63	U	Aveg.	288,75	903,3333	79,25
			i	StDev	23,36713	63,13382	49,43706
312	812	157	U	CoV	0,080925	0,06989	0,623811
312	937	34	U				
312	906	64	U				
312	906	64	U				
281	937	37	U				
281	937	37	U				
281	906	65	U				
281	906	65	U				
281	906	65	U				
250	875	106	U				
250	781	194	U				
0	0	1014	i				

The 10 averages for the 10 /ʉ/-imitations and the 10 /hʉs/-imitations were grouped in single tables to observe how the CoVs have changed between the start test and the end test. These tables can be observed in Chapter 7.5 The Analysis of the coefficients of variance.

6.1 The Test results of the treatment group

The test data of the treatment group regarding the pronunciation of the 10 individual /ʉ/-vowels appears in Appendix A. The first 4 columns display the 10 individual average hits in the start test, the following 4 columns display the average hits in the 4th exercise and the last 4 columns display the corresponding average hits in the end test.

The data regarding the treatment group's pronunciation of the 10 individual vowels in the /hʉs/-words appears in Appendix B. The first 4 columns display the 10 individual average hits in the start test, the following 4 columns display the average hits in the 4th exercise and the last 4 columns display the corresponding average hits in the end test.

The visualization of the sustained /ʉ/-hits appear in Appendix E in the form of 30 diagrams. There are 3 diagrams for each of the 10 test subjects: the start test diagram, the 4th exercise diagram and the end test diagram. The 30 /hʉs/-imitation diagrams of the treatment group appear in Appendix F.

6.2 The Results of the control group

The reference data of the control group regarding the pronunciation of the 10 individual /ʉ/-vowels appears in Appendix C. The first 4 columns display the 10 individual average hits in the start test, the following 4 columns display the average hits in the 4th exercise and the last 4 columns display the corresponding average hits in the end test.

The data regarding the control group's pronunciation of the 10 individual vowels in the /hʉs/-words appears in Appendix D. The first 4 columns display the 10 individual average hits in the start test, the following 4 columns display the average hits in the 4th exercise and the last 4 columns display the corresponding average hits in the end test.

The visualization of the sustained /ʉ/-hits appear in Appendix G in the form of 30 diagrams. There are three diagrams for each of the 10 reference/control subjects: the start test diagram, the 4th exercise diagram and the end test diagram. The 30 diagrams of the /hʉs/-imitations of the control group appear in Appendix H.

7. The Analysis of the results

The test subjects were selected randomly when forming the groups. Before the start test, the subjects were requested to fill in a form in the laboratory regarding their name, age and school background. (The form is presented in its original Finnish form in Appendix I). Each of the 4 visits to the testing laboratory was also marked in the form to keep track of the progress of the tests. It was verified beforehand that all the test subjects had normal hearing, speech production and vision and that they had no Swedish-speaking background other than that learned at school. The only other factor that was monitored in the selection was that the number of males and females was kept constant in both groups (4 females and 6 males). This particular ratio was formed due to the fact that 12 males and 8 females enrolled for the tests. Out of these 20 people, 2 never took part in the test and replacement test subjects were therefore recruited to fill in the gaps. In one case, the test was performed incorrectly and a replacement test subject was brought in. In another case it was revealed that one test subject had a Swedish-speaking background after all and a replacement test subject had to be brought in as well.

The reason for selecting both male and female participants in both groups, despite the known difference in the formant values between the sexes, came out of necessity due to the limitation set by the number of individuals consenting to take part in the tests. A rule of using only males or only females would have drastically decreased the sizes of the groups.

7.1 The General background of the subjects

The test subjects were all between 19 and 30 years of age, with the average age being 22,35 years.

Table 6 The average age of the test subjects

	Age (in years):
Average age:	22,35
Average treatment subject's age:	21,9
Average control subject's age:	22,8
Average Male subject's age:	22,8
Average Female subject's age:	21,6
Average treatment subject's age (Male):	22,8
Average treatment subject's age (Female):	20,5
Average control subject's age (Male):	22,8
Average control subject's age (Female):	22,75

The level of the test subjects' school background regarding Swedish was also inquired, but only after the tests. This was to ensure that the test subjects would not feel uncomfortable about revealing their school grades before the tests. The goal was to make the test situations as relaxed and stress-free as possible to ensure that any tension would not interfere with the results (Eriksson 2004: 29).

Table 7 illustrates the time period between the test subjects' graduation from high school and their participation in the tests. This information was not applied in the analysis of the results.

Table 7 The years since graduation from high school

	Years since graduation:
All	3,2
Treatment	2,8
Control	3,6

Tables 8 and 9 illustrate the subjects' school and graduation grades regarding Swedish. The differences here are arbitrary and their effect on the results can be deemed negligible. Table 10 illustrates the geographical distribution of the test subjects' graduation cities across the two groups. This information simply informs us where the test subjects learned their school Swedish. The information was not applied to the analysis of the results.

Table 8 The test subjects' Swedish graduation grades

	Treatment group:	Control Group:	Total:
B Swedish A Swedish	10 0	8 2	18 2 20
Swedish Grad. Grade	Treatment group:	Control Group:	Total:
L	1	1	2
E	1	1	2
M	4	1	5
C	2	2	4
B	2	3	5
A			

Table 9 The test subjects' average Swedish grades

	Aveg. Mean grade of Swedish courses:	StDev:	CoV
Treatment	7,95	1,257201478	0,158139
Control	7,28	1,845595357	0,253516
All	7,615	1,57489348	0,206815

Table 10 The geographical location of the test subjects' graduation

Graduation City:	All:	Treatment group:	Control group:
Turku	5	2	3
Salo	3	2	1
Loimaa	2	2	-
Tampere	2	1	1
Forssa	1	-	1
Halikko	1	-	1
Helsinki	1	1	-
Kotka	1	-	1
Kuortane	1	1	-
Raisio	1	-	1
Rauma	1	-	1
Vaasa	1	1	-
Total:	20	10	10

7.2 The Kolmogorov-Smirnov analysis

The data of the treatment group and the control group were analysed using a statistics software, SPSS (version 16). The software is available at www.spss.com (SPSS Inc. 2008). The first procedure was to determine if the data of the groups had normal distribution. This is crucial in defining which analysis method should be used. The data of the groups was analysed using the Kolmogorov-Smirnov test (Nummenmaa 2004: 143). The result was that the data did not have normal distribution except in three cases. The data of the treatment group and the control group regarding the F1 and F2 values in the start test of the /hʉs/-imitations had normal distribution and also the F1 values for the /hʉs/-imitations in the end had normal distribution. In all the other cases, the data did not have normal distribution, thus leading to the abandonment of the ANOVA analysis and to selection of the Mann-Whitney test (Nummenmaa 2004: 250). The Mann-Whitney analysis showed that there was a statistically significant difference between the treatment group and the control group in the end test regarding the sustained /ʉ/-imitations and the /hʉs/-imitations.

Table 11 The Kolmogorov-Smirnov analysis of the test data from SPSS (Version 16)

Test Statistics													
		U START F1	U START F2	U START Eucl. Dist	U END F1	U END F2	U END Eucl. Dist	HUs START F1	HUs START F2	HUs START Eucl. Dist	HUs END F1	HUs END F2	HUs END Eucl. Dist
N		194	194	194	200	200	200	194	194	194	200	200	200
Normal Parametersa	Mean	325,13	1138	250,28	330,73	1109,7	207,32	342,84	1202,4	271,76	340,69	1220	278,218
	Std. Deviation	53,913	264,11	221,97	58,777	180,27	147,68	50,771	190,76	166,98	52,561	187,07	174,13496
Most Extreme Differences	Absolute	0,1532	0,1744	0,1814	0,1482	0,1054	0,1562	0,0834	0,0826	0,109	0,0953	0,1047	0,1218249
	Positive	0,1532	0,1744	0,1814	0,1482	0,1054	0,1562	0,0834	0,0826	0,109	0,0953	0,1047	0,1218249
	Negative	-0,082	-0,089	-0,158	-0,085	-0,043	-0,12	-0,045	-0,044	-0,086	-0,066	-0,063	-0,0992288
Kolmogorov-Smirnov Z		2,1337	2,4291	2,5265	2,0952	1,4913	2,2085	1,1621	1,1501	1,5176	1,3484	1,4812	1,7228641
Asymp. Sig. (2-tailed)		0,0002	1E-05	6E-06	0,0003	0,0234	0,0001	0,1342	0,1419	0,02	0,0527	0,0248	0,0052824
a. Test distribution is Normal.													

As it can be seen in Table 11, there is a statistically significant difference between the data and normal distribution in nine of the twelve parameters. The p-value (Asymp. Sig. (2-tailed) is smaller than 5% with:

/U/ START F1
/U/ START F2
/U/ START Eucl. Dist.
/U/ END F1
/U/ END F2
/U/ Eucl. Dist.
/hUs/ START Eucl. Dist
/hUs/ END F2
/hUs/ END Eucl. Dist

This means that an ANOVA analysis (or a t-test) cannot be used and that an alternative statistical method must be used instead. A Mann-Whitney U-test was used, which is the non-parametric equivalent of the t-test.

7.3 The Mann-Whitney U-test

The data had to be analysed using a non-parametric equivalent of the t-test, the Mann-Whitney U-test to see is there was a statistically significant difference between the groups in the start test and in the end test. The analysis was performed with SPSS (version 16).

Table 12 The Mean Rank result of the Mann-Whitney U-test from SPSS (Version 16)

Ranks

	TestRef	N	Mean Rank	Sum of Ranks
U START F1	1	96	106,05	10180,50
	2	98	89,13	8734,50
	Total	194		
U START F2	1	96	100,61	9658,50
	2	98	94,45	9256,50
	Total	194		
U START Eucl. Dist.	1	96	92,98	8926,00
	2	98	101,93	9989,00
	Total	194		
U END F1	1	100	99,18	9918,50
	2	100	101,82	10181,50
	Total	200		
U END F2	1	100	90,09	9009,00
	2	100	110,91	11091,00
	Total	200		
U END Eucl. Dist	1	100	89,39	8939,00
	2	100	111,61	11161,00
	Total	200		
hUs START F1	1	99	94,45	9350,50
	2	95	100,68	9564,50
	Total	194		
hUs START F2	1	99	99,78	9878,50
	2	95	95,12	9036,50

	Total	194		
hUs START Eucl. Dist.	1	99	96,16	9519,50
	2	95	98,90	9395,50
	Total	194		
hUs END F1	1	100	92,32	9232,50
	2	100	108,68	10867,50
	Total	200		
hUs END F2	1	100	90,01	9001,00
	2	100	110,99	11099,00
	Total	200		
hUs END Eucl. Dist.	1	100	89,00	8900,00
	2	100	112,00	11200,00
	Total	200		

Table 13 The Mann-Whitney U-test from SPSS (Version 16)

Test Statistics												
	U START F1	U START F2	U START Eucl. Dist	U END F1	U END F2	U END Eucl. Dist	hUs START F1	hUs START F2	hUs START Eucl. Diist	hUs END F1	hUs END F2	hUs END Eucl Dist
Mann-Whitney U	3883,5	4405,5	4270	4868,5	3959	3889	4400,5	4476,5	4569,5	4182,5	3951	3850
Wilcoxon W	8734,5	9256,5	8926	9918,5	9009	8939	9350,5	9036,5	9519,5	9232,5	9001	8900
Z	-2,1	-0,763	-1,11	-0,321	-2,544	-2,715	-0,773	-0,578	-0,34	-1,998	-2,563	-2,81
Asymp. Sig. (2-tailed)	0,0357	0,4452	0,267	0,748	0,011	0,0066	0,4398	0,5632	0,7337	0,0458	0,0104	0,005
a. Grouping Variable: Treatment/ Control												

As it can be seen in Table 13, there is a statistically significant difference between the groups in 6 cases, 5 of which are end test situations. Asymp. Sig (2-tailed) is denoted in the text as p.

E.g. In the start test, there is a statistically significant difference between the groups regarding the sustained /ʉ/-imitations with regard to F1, but there is no difference between the groups in the end. In Table 12 it can be seen that in the start test the treatment group's mean rank is higher (106,05) than the control group's (89,13), but the situation is reversed in the end test (Treatment group's Mean rank = 99,18, control group's Mean rank = 101,82). This means that in the start test the treatment group's values for F1 were generally higher than those of the control group's, but the situation was reversed in the end.

In the start test, there is no statistically significant difference between the groups regarding the sustained /ʉ/-imitations with regard to F2 (p = 0,4452), but there is a difference between the groups in the end (p = 0,011). The treatment group's mean rank is lower in the end than that of the control group's (see Table 12). This means that the treatment group's values for this parameter (F2) were generally smaller in the end than those of the control group's.

In the start test, there is no statistically significant difference between the groups regarding the sustained /ʉ/-imitations with regard to the Euclidean distance (p = 0,267), but there is a difference between the groups in the end (p = 0,0066). The treatment group's mean rank is lower in the end than that of the control group's (see Table 12). This means that the treatment group's values for this parameter were generally smaller in the end than those of the control group's. More importantly it means that the treatment group's /ʉ/-hits were closer to the reference point in the end test than those of the control group.

In the start test, there is no statistically significant difference between the groups regarding the /hʉs/-imitations with regard to F1 (p = 0,4398), but there is a difference between the groups in the end (p = 0,0458). The treatment group's mean rank is lower in the end than that of the control group's (see Table 12). This means that the treatment group's values for this parameter were generally smaller in the end than those of the control group's.

In the start test, there is no statistically significant difference between the groups regarding the /hʉs/-imitations with regard to F2 (p = 0,5632), but there is a difference between the groups in the end (p = 0,0104). The treatment group's mean rank is lower in the end than that of the control group's (see Table 12). This means that the treatment group's values for this parameter were generally smaller in the end than those of the control group's.

In the start test, there is no statistically significant difference between the groups regarding the /hʉs/-imitations with regard to the Euclidean distance (p = 0,7337), but there is a difference between the groups in the end (p = 0,005). The treatment group's mean rank is lower in the end than that of the control group's (see Table 12). This means that the treatment group's values for this parameter were generally smaller in the end than those of the control group's. More importantly it means that the treatment group's /hʉs/-hits were again closer to the reference point in the end test than those of the control group.

All this shows that there is a statistically significant difference between the groups in the end when there was virtually none in the beginning. This means that either group has benefited from the tests. The next step is to examine the data more closely to determine which group fared better.

7.4. The Wilcoxon signed ranks test within the treatment and control groups

Before the analysis of the coefficients of variance, the Wilcoxon signed ranks test was performed on both groups separately to see if there was a statistically significant difference within the groups between the start test and the end test (Nummenmaa 2004: 253). The analysis was performed with SPSS (version 16).

Table 14 The Wilcoxon test: Euclidean distance: Treatment group: start test and end test

Test Statistics		
	/ʉ/ The Euclidean distance in the start test and in the end test	/hʉs/ The Euclidean distance in the start test and in the end test
Z	-1,304588959	-0,93889
Asymp. Sig. (2-tailed)	0,19203285	0,347788
a. Based on positive ranks.		
b. Based on negative ranks.		
c. Wilcoxon Signed Ranks Test		

Table 15 The Wilcoxon test: Euclidean distance: Control group: start test and end test

Test Statistics		
	/ʉ/ The Euclidean distance in the start test and in the end test	/hʉs/ The Euclidean distance in the start test and in the end test
Z	-1,247340078	-1,10242
Asymp. Sig. (2-tailed)	0,212272828	0,270278
a. Based on positive ranks.		
b. Based on negative ranks.		
c. Wilcoxon Signed Ranks Test		

As Table 14 shows, p is well over 0,05 with both the /ʉ/-vowel and the /hʉs/-word of the treatment group (tr = treatment, $p_{tr,u} = 0{,}192$ and $p_{tr,hus} = 0{,}347$). This shows that within the treatment group there was no statistically significant difference between the Euclidean distance values in the end test and in the start test.

The same test was performed on the control group and the results, which appear in Table 15, illustrated the same outcome (cont = control, $p_{cont,u} = 0{,}212$ and $p_{cont,hus} = 0{,}270$). With the control group, there was no statistically significant difference between the Euclidean distance values in the end test and in the start test

Table 16 The Wilcoxon test: Treatment group: F1 and F2

Test Statistics				
	/ʉ/ F1 in the start test and in the end test	/ʉ/ F2 in the start test and in the end test	/hʉs/ F1 in the start test and in the end test	/hʉs/ F2 in the start test and in the end test
Z	-1,74252	-0,40928	-0,64222	-0,80626
Asymp. Sig. (2-tailed)	0,081417	0,682332	0,520733	0,420095
a. Based on positive ranks.				
b. Based on negative ranks.				
c. Wilcoxon Signed Ranks Test				

Table 17 The Wilcoxon test: Control group: F1 and F2

Test Statistics				
	/ʉ/ F1 in the start test and in the end test	/ʉ/ F2 in the start test and in the end test	/hʉs/ F1 in the start test and in the end test	/hʉs/ F2 in the start test and in the end test
Z	-3,60922	-0,16012	-0,09651	-1,40308
Asymp. Sig. (2-tailed)	0,000307	0,872785	0,923117	0,160592
a. Based on positive ranks.				
b. Based on negative ranks.				
c. Wilcoxon Signed Ranks Test				

Table 16 illustrates that there is no statistically significant difference between the start test and the end test within the treatment group with regard to F1 and F2. The result applies for both the sustained /ʉ/-vowel (tr = treatment, $p_{tr,U,F1} = 0{,}081$ and $p_{tr,hus,F2} = 0{,}682$) and the /hʉs/-word ($p_{tr,hUs,F1} = 0{,}521$ and $p_{tr,hus,F2} = 0{,}420$). The p-values are well over 0,05.

Table 17 illustrates that there is a statistically significant difference in the control group between the start test and the end test with regard to one parameter: F1 for /ʉ/, $p_{cont,U,F1} = 0{,}000307$. With F2 for /ʉ/ and with F1 and F2 for /hʉs/ there is no statistically significant difference between the start and the end (cont = control, $p_{cont,U,F2} = 0{,}873$) and (cont = control, $p_{cont,hUs,F1} = 0{,}923$ and $p_{cont,hUs,F2} = 0{,}161$). This is examined more closely in Chapter 7.6.

7.5 The Analysis of the coefficients of variance

The hypothesis was that the treatment group's CoVs for the F1, F2 and the Euclidean distance would be smaller in the end test when compared to the corresponding values from the start test. The hypothesis also stated that the treatment group's CoV-values would have decreased in a greater number of cases when compared to the corresponding performance of the control group.

When observing the diagrams of the treatment group members' average hits of the sustained /ʉ/-imitations in Appendix E, it can be seen that the standard deviation has clearly decreased in 8 cases. In these cases, the group of 10 average points in the diagrams is visibly tighter in the end test when compared to the start situation. When examining the /hʉs/-imitations, the number is only 5 (see Appendix F). When observing the control group's /ʉ/-imitations, it can be seen that the deviation of the average hits has decreased in only 6 cases (see Appendix G). The corresponding number with the /hʉs/-imitations is only 4 (see Appendix H).

Table 18 Treatment group: CoVs of the /ʉ/-imitations (subjects 1-10)

START	CoV			4th Ex.	CoV			END	CoV		
/U/	F1	F2	Eucl. Dist.	/U/	F1	F2	Eucl. Dist.	/U/	F1	F2	Eucl. Dist.
1	0,20619	0,20236	0,41304	1	0,03648	0,08491	0,24354	1	0,03102	0,07954	0,17081
2	0,03108	0,10042	0,2964	2	0,02591	0,10076	0,62934	2	0,02031	0,06194	0,38229
3	0,13869	0,203	0,76975	3	0,03353	0,05766	0,453	3	0,03595	0,09865	0,32954
4	0,08098	0,0506	0,37406	4	0,04067	0,05766	0,23628	4	0,03131	0,0674	0,41224
5	0,0849	0,0513	0,44021	5	0,05494	0,02426	0,22407	5	0,03917	0,04525	0,37768
6	0,12259	0,10491	0,29985	6	0,04317	0,02921	0,07673	6	0,07226	0,0441	0,11896
7	0,1491	0,10173	0,71669	7	0,09118	0,07491	0,34235	7	0,11204	0,05549	0,33737
8	0,03295	0,05244	0,36592	8	0,01664	0,04643	0,42575	8	0,03579	0,06652	0,53248
9	0,06841	0,06502	0,4557	9	0,0276	0,04572	0,24582	9	0,01965	0,06452	0,54431
10	0,09385	0,07788	0,31778	10	0,04995	0,03532	0,23603	10	0,04329	0,03167	0,17867
Aveg:	0,10087	0,10097	0,44494	Aveg:	0,04201	0,05568	0,31129	Aveg:	0,04408	0,06151	0,33844
StDev:	0,05432	0,05763	0,16702	StDev:	0,02072	0,02484	0,15567	StDev:	0,02802	0,01906	0,14561
CoV:	0,53852	0,5708	0,37537	CoV:	0,49327	0,44604	0,50007	CoV:	0,63579	0,30981	0,43025

Table 19 Treatment group: Average values of the /ʉ/-imitations (subjects 1-10)

/U/	START			/U/	END		
Test. Subj.	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	Test. Subj.	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	349,2972	1654,234	709,3417	1	389,0649	828,1892	196,8105
2	283,2352	1359,643	402,3571	2	289,0174	1153,897	185,931
3	296,135	913,0117	157,9517	3	284,4459	969,604	111,2623
4	360,4722	997,661	92,93407	4	298,5834	982,8486	90,06393
5	297,3611	1047,407	93,68519	5	296,1459	1072,979	115,6882
6	417,9714	1135,942	256,916	6	446,8379	1478,418	533,8124
7	359,0938	1121,75	247,9375	7	368,6167	1115,705	184,0631
8	307,4144	1013,889	75,95433	8	301,4863	1063,739	121,224
9	303,1182	954,9028	77,49744	9	271,8997	999,7639	111,537
10	367,3212	1188,21	280,4358	10	333,1116	1121,979	173,7044
Aveg:	334,142	1138,665	239,5011	Aveg:	327,921	1078,712	182,4097
StDev	43,12706	222,5754	197,6837	StDev	56,44465	169,6416	129,3532
CoV	0,129068	0,19547	0,825398	CoV	0,172129	0,157263	0,709136

When examining the treatment group's CoVs for F1 from the sustained /ʉ/-imitations, it can be seen in Table 18 that within the treatment group, the CoV value is smaller in the end test with 9 test subjects out of 10 (90%) when compared to the start test. With F2, the CoV value is smaller in the end in 8 cases (80%). The CoV for the Euclidean distances in the case of the sustained /ʉ/-imitations was smaller in the end in 6 cases (60%). Between the start test and the 4th exercise, the CoV for the F1 value is smaller in the 4th exercise in 10 cases (100%) and with F2 the CoV value is smaller 8 cases (80%). Here it is important to note that the actual average Euclidean distance of the /ʉ/-imitations had grown smaller in the end test with 6 test subjects (see Table 19).

Table 20 Treatment group: CoVs of the /hʉs/-imitations (subjects 1-10)

START:	CoV			4th Ex.	CoV			END	CoV		
/U/	F1	F2	Eucl. Dist.	/U/	F1	F2	Eucl. Dist.	/U/	F1	F2	Eucl. Dist.
1	0,03244	0,05475	0,20423	1	0,03808	0,05763	0,19926	1	0,06453	0,09195	0,29704
2	0,01969	0,14856	0,55241	2	0,02333	0,08959	0,4017	2	0,03084	0,04675	0,21006
3	0,03828	0,06297	0,439	3	0,03389	0,04653	0,29072	3	0,0897	0,15847	0,78298
4	0,06638	0,05554	0,25274	4	0,03477	0,0515	0,29472	4	0,04351	0,09371	0,55364
5	0,16973	0,08297	0,74625	5	0,04895	0,03959	0,23574	5	0,05611	0,03133	0,20453
6	0,04955	0,07652	0,19086	6	0,0461	0,02993	0,0779	6	0,03964	0,04026	0,10156
7	0,07271	0,05032	0,18534	7	0,10756	0,05219	0,18963	7	0,09412	0,05706	0,22792
8	0,06258	0,07649	0,474	8	0,04042	0,06188	0,37878	8	0,05044	0,04009	0,27405
9	0,06202	0,05899	0,23583	9	0,06577	0,03071	0,18657	9	0,0391	0,06661	0,37133
10	0,06334	0,04893	0,17416	10	0,05778	0,03689	0,15344	10	0,05147	0,02808	0,1115
Aveg:	0,06367	0,0716	0,34548	Aveg:	0,04966	0,04964	0,24084	Aveg:	0,05595	0,06543	0,31346
StDev:	0,04095	0,02952	0,19669	StDev:	0,02374	0,01776	0,101	StDev:	0,02123	0,04002	0,21052
CoV:	0,64321	0,41226	0,56932	CoV:	0,47806	0,35785	0,41935	CoV:	0,37943	0,61161	0,67161

Table 21 Treatment group: Average values of the /hʉs/-imitations (subjects 1-10)

/hUs/	START			/hUs/	END		
Test. Subj.	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	Test. Subj.	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	375,3269	1154,393	246,261	1	398,3889	1101,95	223,6208
2	284,1898	1299,493	339,0666	2	286,5977	1052,167	120,7249
3	312,93	1093,947	135,8527	3	315,3617	1204,117	240,3859
4	360,9301	1217,609	260,2899	4	332,331	1123,524	176,7554
5	283,0852	1039,474	105,3648	5	297,9617	1133,627	168,2555
6	418,6565	1311,244	405,1198	6	420,9601	1511,326	562,8936
7	322,4767	1315,987	355,9567	7	334,2119	1272,811	317,0248
8	331,6471	1126,653	171,7794	8	306,5994	1099,275	143,8248
9	340,4366	1098,891	160,8225	9	277,5422	1103,157	160,6491
10	342,9369	1309,227	347,3216	10	355,5684	1250,24	291,7595
Aveg:	337,2616	1196,692	252,7835	Aveg:	332,5523	1185,219	240,5894
StDev	41,1912	106,7484	105,8068	StDev	47,12448	134,9188	129,9816
CoV	0,122134	0,089203	0,418567	CoV	0,141705	0,113834	0,540263

When examining the treatment group's CoVs for F1 from the /hʉs/-imitations, it can be seen in Table 20 that within the treatment group, the CoV value is smaller in the end test in only 6 cases (60%) when compared to the start test. With F2, the CoV value is smaller in the end in only 5 cases (50%). The CoV for the Euclidean distances in the case of the /hʉs/-imitations was smaller in the end in 5 cases (50%). Between the start test and the 4th exercise, the CoV for the F1 value is smaller in the 4th exercise in 6 cases (60%) and with F2 the CoV value is smaller 8 cases (80%). Here it is important to note that the actual average Euclidean distance of the /hʉs/-imitations had grown smaller in the end test with 7 test subjects (see Table 21).

Table 22 Control group: CoVs of the /ʉ/-imitations (subjects 11-20)

START:	CoV			4th Ex.	CoV			END	CoV		
/U/	F1	F2	Eucl. Dist.	/U/	F1	F2	Eucl. Dist.	/U/	F1	F2	Eucl. Dist.
11	0,05998	0,04133	0,10169	11	0,0377	0,09417	0,30043	11	0,03662	0,10277	0,35
12	0,04194	0,1106	0,78285	12	0,03754	0,10632	0,36743	12	0,03772	0,08154	0,49427
13	0,10528	0,0528	0,33433	13	0,06597	0,08829	0,37229	13	0,11948	0,08554	0,40058
14	0,03765	0,17126	0,76682	14	0,03544	0,12902	0,34061	14	0,05461	0,13129	0,45051
15	0,02115	0,30138	0,79942	15	0,06833	0,06551	0,31651	15	0,07493	0,07545	0,45341
16	0,04099	0,09103	0,48417	16	0,04204	0,10021	0,51459	16	0,03436	0,04166	0,37696
17	0,03255	0,07833	0,3273	17	0,03302	0,20483	0,99397	17	0,0659	0,16673	0,66218
18	0,1627	0,08814	0,20658	18	0,17858	0,20705	0,59727	18	0,19425	0,15338	0,43972
19	0,12886	0,14405	0,50449	19	0,0399	0,0455	0,16649	19	0,05756	0,06194	0,24473
20	0,03711	0,11291	0,65944	20	0,05835	0,05511	0,28882	20	0,03446	0,03336	0,17877
Aveg:	0,06682	0,11918	0,49671	Aveg:	0,05969	0,1096	0,42584	Aveg:	0,07099	0,09337	0,40511
StDev:	0,04813	0,07493	0,25132	StDev:	0,04375	0,05652	0,23269	StDev:	0,05055	0,04509	0,13351
CoV:	0,72024	0,62872	0,50597	CoV:	0,733	0,51572	0,54642	CoV:	0,71207	0,48289	0,32957

Table 23 Control group: Average values of the /ʉ/-imitations (subjects 11-20)

/U/	START			/U/	END		
Cont. Subj.	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	Cont. Subj.	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
11	315,6741	1632,954	664,4056	11	340,1336	1355,039	390,5269
12	274,5583	1045,894	162,1967	12	257,1461	1038,87	134,3812
13	383,1139	941,9931	121,1446	13	376,6705	1025,088	149,9591
14	264,9169	986,8148	183,9881	14	277,418	1005,378	252,8459
15	329,5903	1237,14	355,7232	15	351,3563	1046,428	120,1608
16	300,7838	913,6444	116,7483	16	286,9626	1032,316	76,83642
17	262,9683	1132,318	205,33	17	285,9072	1108,451	237,0922
18	339,9774	1544,783	600,2009	18	383,1452	1426,478	478,8226
19	390,4838	1020,016	181,7941	19	430,8079	1180,369	256,4337
20	306,1689	1047,753	125,3316	20	345,8847	1188,547	225,1576
Aveg:	316,8236	1150,331	271,6863	Aveg:	333,5432	1140,696	232,2216
StDev	45,05858	249,6556	202,6117	StDev	55,49703	147,2695	124,7219
CoV	0,14222	0,217029	0,745756	CoV	0,166386	0,129105	0,537081

When examining the control group's CoVs for F1 from the sustained /ʉ/-imitations in Table 22, it can be seen that within the control group, the CoV value is smaller in the end test in only 5 cases (50%). With F2 the CoV value is smaller in the end in only 6 cases (60%). The CoV for the Euclidean distances in the case of the sustained /ʉ/-imitations was smaller in the end in 6 cases (60%). Between the start test and the 4th exercise, the CoV for the F1 value is smaller in the 4th exercise in 5 cases (50%) and with F2 the CoV value is smaller in 5 cases (50%). Here it is important to note that the actual average Euclidean distance of the /ʉ/-imitations had grown smaller in the end test with only 5 control subjects (see Table 23).

Table 24 Control group: CoVs of the /hʉs/-imitations (subjects 11-20)

START	CoV			4th Ex.	CoV			END	CoV		
/U/	F1	F2	Eucl. Dist.	/U/	F1	F2	Eucl. Dist.	/U/	F1	F2	Eucl. Dist.
11	0,06566	0,06189	0,20441	11	0,05706	0,07404	0,21215	11	0,05165	0,07115	0,22404
12	0,06575	0,06505	0,37267	12	0,05522	0,07532	0,58538	12	0,0619	0,05571	0,38286
13	0,02721	0,09193	0,26655	13	0,11069	0,08429	0,22156	13	0,11772	0,11232	0,31413
14	0,02445	0,06274	0,32512	14	0,09423	0,10973	0,36808	14	0,04799	0,10335	0,34565
15	0,03596	0,15777	0,52567	15	0,05447	0,05935	0,30266	15	0,04084	0,04104	0,21827
16	0,04243	0,07291	0,31915	16	0,05806	0,13822	0,54549	16	0,04726	0,05256	0,39968
17	0,07238	0,16104	0,75807	17	0,05435	0,09352	0,47346	17	0,11566	0,1605	0,72361
18	0,14103	0,10735	0,27523	18	0,0903	0,05543	0,13277	18	0,0531	0,05799	0,14827
19	0,05905	0,03635	0,08572	19	0,02445	0,05553	0,17041	19	0,02548	0,04139	0,1494
20	0,04579	0,07436	0,3281	20	0,05832	0,03474	0,13874	20	0,07394	0,03998	0,14713
Aveg:	0,05797	0,08914	0,34607	Aveg:	0,06572	0,07802	0,31507	Aveg:	0,06356	0,0736	0,3053
StDev:	0,03362	0,04151	0,1836	StDev:	0,02516	0,03013	0,16958	StDev:	0,0307	0,03969	0,17631
CoV:	0,57991	0,46565	0,53053	CoV:	0,38287	0,38619	0,53821	CoV:	0,48308	0,53925	0,5775

Table 25 Control group: Average values of the /hʉs/-imitations (subjects 11-20)

/hUs/	START			/hUs/	END		
Cont. Subj.	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	Cont. Subj.	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
11	383,0139	1232,739	304,67	11	350,2864	1397,702	433,798
12	297,4071	1001,517	104,0359	12	289,6581	1064,062	120,1101
13	414,4087	969,1152	174,3019	13	386,5383	1395,987	450,2083
14	305,335	1199,26	231,449	14	285,0839	971,8123	142,8769
15	345,7741	1357,268	395,6455	15	384,8755	1147,696	203,5475
16	336,0042	933,7787	100,3293	16	308,2911	1099,502	143,7512
17	270,8944	1213,989	251,4611	17	307,4079	1234,817	273,4688
18	341,8611	1580,704	617,2037	18	388,8563	1629,493	677,4816
19	447,1939	1485,148	543,0473	19	446,4269	1266,844	337,8102
20	307,3532	1132,822	195,0977	20	340,7915	1339,814	375,413
Aveg:	344,9246	1210,634	291,7241	Aveg:	348,8216	1254,773	315,8466
StDev	55,32214	215,8203	176,4795	StDev	52,53494	194,5287	176,078
CoV	0,160389	0,17827	0,604953	CoV	0,150607	0,155031	0,557479

When examining the control group's CoVs for F1 from the /hʉs/-imitations, it can be seen in Table 24 that the CoV value is smaller in the end test in only 4 cases (40%). With F2 the CoV value is smaller in the end in 6 cases (60%). The CoV for the Euclidean distances in the case of the /hʉs/-imitations was smaller in the end in only 3 cases (30%). Between the start test and the 4th exercise, the CoV for the F1 value is smaller in the 4th exercise in 5 cases (50%) and with F2 the CoV value is smaller in 5 cases (50%). Here it is important to note that the actual average Euclidean distance of the /hʉs/-imitations had grown smaller in the end test with only 3 reference subjects (see Table 25).

Table 26 Percentage of cases where the CoV was smaller in the end test within the group

Treatment group	F1	F2	Eucl. Dist
/U/	90%	80%	60%
/hUs/	60%	50%	50%
Control group	F1	F2	Eucl. Dist
/U/	50%	60%	60%
/hUs/	40%	60%	30%

As Table 26 illustrates, the treatment group's deviation in the end test was smaller in a higher number of cases when compared to the control group. This would indicate that the original hypothesis was correct. However, the final conclusions can be made from Tables 28 and 29.

Table 27 Percentage of cases where the Euclidean distance was smaller in the end test within the group

Treatment group	Eucl. Dist
/U/	60%
/hUs/	70%
Control group	Eucl. Dist
/U/	50%
/hUs/	30%

Table 27 illustrates that a higher number of treatment group members were able to target their hits closer to the reference point in the end test when compared to the start test.

7.6 The Key findings regarding the effects

This chapter draws together the most relevant key findings regarding the effects of The Vowel Game's real-time visual feedback.

Table 28 The average /ʉ/-data of the treatment and control groups

	/U/	START			/U/	END		
Treatment	n=96	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	n=100	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
	Aveg.	333,8474	1134,598	235,9501	Aveg.	327,921	1078,712	182,4097
	StDev	55,61968	250,4295	220,289	StDev	56,78333	173,6343	133,0896
	CoV	0,166602	0,220721	0,933626	CoV	0,173162	0,160964	0,729619
	/U/	START			/U/	END		
Control	n=98	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	n=100	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
	Aveg:	316,599	1141,381	264,3268	Aveg:	333,5432	1140,696	232,2216
	StDev:	51,03771	278,1025	223,8393	StDev:	60,86116	182,3057	157,7047
	CoV:	0,161206	0,243654	0,846828	CoV:	0,182469	0,15982	0,679113
	p	0,0357	0,4452	0,267	p	0,748	0,011	0,0066

Table 28 displays the averages of 100 sustained /ʉ/-imitations of both groups in the start test and in the end test (notice the few excluded outliers in the n-values).

As it can be seen in Table 28, F1 is the only parameter with which there is no statistically significant difference between the groups in the end test, though there is a difference in the start (based on the p-values obtained from the Mann-Whitney U-test). With F2 and the Euclidean distance, there is a statistically significant difference in the end.

With /ʉ/, the average Euclidean distance of the treatment group in the end is 182,41 Hz (a change of -22,7% from the start) whereas the control group's distance is 232,22 Hz (a change of –12,1% from the start), with the difference being 49,81 Hz. The difference is to the treatment group's favour (182,41 Hz < 232,22 Hz).

As Table 28 shows, the control group's F1 value is closer to the reference point in the start test, but the situation is reversed in the end test.

Within the treatment group, the average F1 value has changed between the start test and the end test by -5,93 Hz (-1,8 %). As the Wilcoxon-test in Chapter 7.4 revealed, there is no statistically significant difference within the treatment group between the start test and the end test with regard to F1, F2 and the Euclidean distance.

What is noticeable with /ʉ/ is the change in the treatment group's F2 value between the start test and the end test. The treatment group's average F2 in the start is 1134,60Hz and the average F2 in the end is 1078,71 Hz. The change is -55,89 Hz (–4,9%). According J.L. Flanagan's studies in 1955, a perceivable or audible change occurs, when the change in the formant values is over 3,8% (+/-) (Flanagan in Rosner and Pickering 1994: 55).

Despite the fact that the Wilcoxon-test in Chapter 7.4 revealed that there is no statistically significant difference within the treatment group between the start test and the end test across all the parameters, there is still a perceivable change in the treatment group's F2 value for /ʉ/.

Within the control group with /ʉ/, the change between the start test and the end test with F1 is +16,94 Hz (+5,4%). As the Wilcoxon test in Chapter 7.4 revealed, there is a statistically significant difference in the control group between the start test and the end test with regard to F1 in /ʉ/. The control group's F2 value does not change more than –0,06%.

This means that with /ʉ/, the treatment group has not been able to create a perceivable change with regard to F1 (-1,8 %) according to Flanagan's theory, but there is a perceivable effect with regard to F2 (-4,9%).

In the control group's case, the change is perceivable with F1 (+5,4%), but towards the wrong direction (away from 300 Hz, which was the reference point's F1 value).

Table 29 The average /hʉs/-data of the treatment and control groups

	/hUs/	START			/hUs/	END		
Treatment	n=99	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	n=100	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
	Aveg.	337,8088	1198,28	254,2726	Aveg.	332,5523	1185,219	240,5894
	StDev	44,8622	135,0375	129,4832	StDev	48,79908	154,0444	146,1175
	CoV	0,132804	0,112693	0,50923	CoV	0,146741	0,129971	0,607331
	/hUs/	START			/hUs/	END		
Control	n=95	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	n=100	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
	Aveg:	348,0739	1206,597	289,9933	Aveg:	348,8216	1254,773	315,8466
	StDev:	56,03499	235,9327	197,7701	StDev:	55,1199	210,1677	191,6808
	CoV:	0,160986	0,195536	0,681982	CoV:	0,158017	0,167495	0,606879
	p	0,4398	0,5632	0,7337	p	0,0458	0,0104	0,005

Table 29 displays the averages of 100 /hʉs/-imitations of both groups in the start test and in the end test (notice the few excluded outliers in the n-values). As it can be seen, there is no statistically significant difference between the groups in the start test, but there is a clear statistical difference in the end across all the parameters.

As Table 29 shows, the treatment group’s average Euclidean distance is smaller than the control group’s in the start test. The treatment group’s Euclidean distance has decreased in the end whereas the control group’s distance has increased and has therefore drawn further away from the reference point.

With /hʉs/, the average Euclidean distance of the treatment group in the end is 240,59 Hz (a change of -5,4% from the start) whereas the control group’s distance is 315,85 Hz (a change of +8,9% from the start), with the difference being 75,26 Hz between the groups. This is again to the treatment group’s favour (240,59 Hz < 315,85 Hz).

In the end test with /hʉs/, the difference between the groups with regard to F1 is 16,27 Hz (4,7%) and with F2 the difference is 69,55 Hz (5,5%).

Within the treatment group, the average F1 value has changed between the start test and the end test by -5,26 Hz (-1,6 %). With F2 the change is -13,1 Hz (-1,1 %). This corresponds with the Wilcoxon-test in Chapter 7.4, which revealed that there was no statistically significant difference in the treatment group between the start test and the end test.

Within the control group the changes between the start test and the end test are +0,75 Hz (+0,2%) with F1 and +48,18 Hz (+4%) with F2.

This means that with /hʉs/, the treatment group has not been able to create a perceivable change between the start test and the end test with regard to F1 and F2 according to Flanagan's theory and the Wilcoxon-test.

In the control group's case, the change is only just perceivable with F2, but towards the wrong direction (away from 969 Hz, which was the reference point's F2 value). Also, the control group's Euclidean distance has increased towards the wrong direction (+8,9%)

The statistics of the entire data with regard to minimum and maximum values is presented in Table 30.

Table 30 The statistics of the entire data

		U START F1	U START F2	U START Eucl. Dist	U END F1	U END F2	U END Eucl. Dist	HUs START F1	HUs START F2	HUs START Eucl. Dist	HUs END F1	HUs END F2	HUs END Eucl. Dist
N	Valid	194	194	194	200	200	200	194	194	194	200	200	200
	Missing	6	6	6	0	0	0	6	6	6	0	0	0
	Mean	325,134	1138,02	250,285	330,732	1109,7	207,316	342,836	1202,35	271,765	340,687	1220	278,2179997
	Median	312	1042,22	162,068	312	1088,05	165,994	332,833	1187	243,667	328,885	1173,75	227,8
	Std. Deviation	53,9134	264,107	221,97	58,7771	180,273	147,676	50,7711	190,756	166,984	52,5611	187,069	174,1349596
	Variance	2906,66	69752,5	49270,6	3454,75	32498,2	21808,2	2577,71	36387,7	27883,6	2762,66	34995	30322,98417
	Minimum	250	750	24,6667	250	743,909	19	250	817,333	40	250	843,444	44,8
	Maximum	500	1947,33	979,667	491,75	1968	1000	491,75	1875	906	458,9	1774,8	868,3333333

7.7 The Feedback from the subjects regarding the tests

The test subjects were issued a feedback form after the end test to monitor the quality and efficiency of the actual test setting and the general usability of The Vowel Game. All the 20 subjects filled in the feedback form and the results appear in Tables 31 and 32. The numbers on the rows indicate the number of subjects giving the rating specified at the top of the columns. The feedback form is presented in its original Finnish form in Appendix J. A similar feedback questionnaire was also utilised on children by A-M. Öster in her studies, which evaluated a number of training software (e.g. SpeechViewer) regarding computer-based speech therapy (2006: 166).

Table 31 The Feedback form results

Test quality:	Excellent	Good	Mediocre	Weak	Poor
How would you rate the way the tests were organized?	17	3	-	-	-
How would you rate the tests' interest level?	-	16	4	-	-
How would you rate the compensation? (The cinema ticket)	10	10	-	-	-

Table 32 The Feedback form results 2

Test quality:	I strongly agree	I agree	Not sure	I disagree	I strongly disagree
The tests could be improved / diversified.	1	5	6	7	1
The tests taught me something new.	4	10	5	1	-

The test subjects who utilised The Vowel Game's interface were generally pleased with the application's appearance and usability after they had been initiated to the system.

8. Conclusions

All the analyses in Chapter 7 seem to suggest that The Vowel Game's visual feedback does have a positive effect on the treatment group's pronunciation of the /ʉ/-vowel. The positive development comes across as a progressively more constant pronunciation of the vowels with less deviation in the F1, F2 and Euclidean distance values in the end.

There was a clear statistically significant difference between the groups in the end test when there was none in the start tests (except with /ʉ/ in F1) (see Tables 13, 28 and 29) and the deviations in the treatment group's parameters had decreased in the end test in a higher number of cases when compared to the control group (see Table 26). In addition, the Euclidean distance of the treatment group had decreased in the end test in a higher number of cases when compared to the control group (see Table 27).

Despite the fact that there was no statistically significant difference within the groups between the start test and the end test with regard to F1, F2 and the Euclidean distance (see Chapter 7.4), there was a clear statistically significant difference between the groups in the end with regard to F2 and the Euclidean distance with /ʉ/. The statistically significant difference between the groups was also apparent in the end test with /hʉs/ across all the parameters. The treatment group was closer to the reference point than the control group in both end tests involving both the sustained /ʉ/-vowel and the /hʉs/-word. Generally, all the treatment group's parameter values in the end test with both /ʉ/ and /hʉs/ were closer to the reference point when compared to the control group's end test values (see Tables 28 and 29).

The most prevalent effect within the groups came across in the treatment's group's F2 value for the sustained /ʉ/-vowel between the start test and the end test. The change was 4,9 % towards the correct direction (towards the reference point's F2 value, which was 969 Hz). This, according to J.L. Flanagan's theory, is a perceivable difference (see Chapter 7.6).

In general, the control group had developed as well and learned more constant pronunciation through imitation, but the control group's progress was less apparent when compared to the treatment group. The learning of the control group by imitation alone could be explained by the existence of the mirror neuron system in humans as suggested by Rizzolatti, Craighero and Arbib. Of course, the primary purpose of this study was not to show the effect of the mirror neurons or to prove that their existence caused the learning, but merely to imply that it is one plausible explanation for the learning process.

The use of The Vowel Game does require the guidance and the presence of a phonetics specialist or speech clinician when the goal is to learn new vowel categories. For instance, in the case of the /ʉ/-vowel, The Vowel Game does not offer any information regarding the roundedness of the lips, which is crucial with this particular vowel. This type of information was provided for the test subjects in both groups by the supervisor. Also, as M. Carey stated in his study, the F1/F2-plane alone does not provide information on how the formants should be altered. This requires an external advisor. What appears to be the common denominator in most of the systems presented in Chapter 4 is that they often require the presence of a specialist who guides and instructs the subjects. Many of the applications in Chapter 4 were also deemed quite complex to use, e.g. by M. Carey and D. Healey in their respective evaluations. However, the subjects who used The Vowel Game's interface were generally pleased with the application's appearance and usability after they had been initiated to the system.

As stated earlier, The Vowel Game's game-feature, where the student receives points from hitting any of the prototypes of the vowel categories in the diagram, was excluded from the test version and therefore the game aspect could not be evaluated. In any case, the game aspect in the application is still in its infancy and should be developed further.

In general, the majority of the subjects found the tests of The Vowel Game to be well organized and generally interesting, regardless of whether they were treatment or control subjects. The majority of the subjects also felt they had learned something new in the tests. The evidence of the feedback was presented in Tables 31 and 32.

The tests of The Vowel Game should be continued with larger groups over longer periods of time to gather more conclusive data. What should be taken into account in the future studies is the selection of the test subjects and the difference between the male and the female voices. According to Rosner and Pickering, the female voices feature formants, which are approximately 18,5% higher than those of males. This is largely due to the difference between the vocal tract lengths between the sexes (1994: 54). In the tests of The Vowel Game, the test subjects consisted of both males and females due to the limitation in the number of individuals consenting to participate in the tests, albeit with a compensation in the form of a cinema ticket. The presence of both sexes in a single group obviously caused deviation in itself in the F1, F2 and the Euclidean distance values. This problem could be overcome in the future tests by selecting only males or only females in the groups and comparing groups of the same sex.

The Vowel Game's performance should also be improved. There are a number of issues in the game, which should be addressed in the future versions of the application to improve the game's performance. 1) The gating of the microphone line should be adjusted to prevent the game from picking up undesirable interferences and disturbances such as processor noise, ventilation noise in the room and the test subject's breathing and consequently from plotting them in the vowel diagram as hits. 2) The merging of the formants in the spectral envelope leads to false hits, when the game interprets the two maxima in the spectral envelope as one and then proceeds to select the third maximum as the second. This obviously leads to a false hit. This problem could be overcome by making an adjustment to the game whereby the program should be able to separate at least 5 maxima in the spectral envelope over a bandwidth of e.g. 200-4000 Hz. 3) The problem regarding the effect of a high F0, which leads to the LPC algorithm's weakness of predicting the multiples of F0 as formants instead of predicting the real formants. This problem may be difficult to bypass, but the program could nevertheless be informed if the test subject has a high F0 (>200 Hz). The program could also be developed to concentrate on a specific area in the diagram and to ignore all hits where the parameters do not correspond with the target area's specifications. E.g. If the target area is the /ʉ/-vowel, the F1 values should not exceed 500 Hz and the F2-values should therefore not exceed 2000 Hz. In case the values are exceeded, the hit is ignored.

With these aforementioned improvements, The Vowel Game could have great potential in becoming an efficient tool in teaching new vowel categories for second language learners. Especially foreign students might well benefit from the advantages of the real-time visual feedback. The application could also be utilised in speech therapy involving children with speech impediments or individuals with hearing disabilities. The future will show how the technical challenges of the application can be overcome.

References

Accent Technologies 1997. Accent lab: Pronunciation place [online]. Available: www.accentlab.com (5th May 2008).

Alku P. & Bäckström T. 2004. Linear predictive method for improved spectral modelling of lower frequencies of speech with small prediction orders. *IEEE Transactions on Speech and Audio Processing*. 12(2): 93-99.

Arbib M.A. 2002. The Mirror system, imitation, and the evolution of language. In Nehaniv C. and Dautenhahn K. (eds): 229-280.

Boe L-J., Schwartz J-L. & Vallee Nathalie 1994. The Prediction of vowel systems: Perceptual contrast and stability. In Keller E. (ed.) [1994] 1999: 185-213.

Boersma P. & Weenink D. 2008. Praat: Doing phonetics by computer [online]. Available: www.fon.hum.uva.nl/praat (5th May 2008).

CALI, Inc. & Formavision 2000. Ellis Master pronunciation TM [online]. Available: www.formavision.com/cateur/it070007.html (5th May 2008).

Carey M. 2002. An L1 specific CALL pedagogy for the instruction of pronunciation with Korean learners of English [online]. Available: www.shlrc.mq.edu.au/~michael/carey_phd.pdf (5th May 2008).

Carlson A.B. 1986. *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*. Singapore: McGraw-Hill.

Chiba T. & Kajiyama M. [1941] 1958. *The Vowel: It's Nature and Structure*. Tokyo: Phonetic Society of Japan.

Cook V. 1991. *Second Language Learning and Language Teaching*. London: Edward Arnold.

Deller J. R., Hansen J. H. L. & Proakis J. G. [1993] 2000. *Discrete-Time Processing of Speech Signals*. Wiley-IEEE Press.

Dew D. & Jensen P. J. 1977. *Phonetic Processing: The Dynamics of Speech*. Columbus: Charles M. Merrill Publishing Company.

Dowd A., Smith J. & Wolfe J. 1998. Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time. *Language and Speech* 41: 1-20.

Ellis R. 1985. *Understanding Second Language Acquisition*. Oxford: Oxford University Press.

English Computerized Learning, Inc. 2006. Pronunciation power [online]. Available: www.englishlearning.com (5th May 2008).

Engstrand O. 1999. Swedish. In International Phonetic Association: 140-142.

Eriksson E. 2004. *User-based Design Recommendations for a Computer-based Speech Training System*. Stockholm: Royal Institute of Technology.

Fant G. [1960] 1970. *Acoustic Theory of Speech Production*. Hague: Mouton & Co. N.V., Publishers.

Fant G. 2004. *Speech Acoustics and Phonetics*. Dordrecht: Kluwer Academic Publishers.

Fipse grant database, Cleveland State University 2005. Technology Enhanced Accent Modification for international teaching assistants [online]. Available: www.fipse.aed.org/grantshow.cfm?grantNumber=P116B920809 (5th May 2008).

Fowler C.A. 1977. *Timing Control in Speech Production*. Indiana: Indiana University Linguistics Club.

Frölunda data 2007. Trollerilådan [online]. Available: www.frolundadata.se/index.cgi?cmd=Shop&cat=3&ucat=350&prod=243 (5th May 2008).

Grillner S., Lindblom B., Lubker J. & Persson A. (eds.) 1987. *Speech Motor Control*. (Werner-Gern Symposium Series 36) Pergamon Press.

Hammerly H. 1991. *Fluency and Accuracy*. Clevedon: Multilingual Matters Ltd.

Healey D. 1998. English language institute: Technology tip of the month: June 1998: Pronunciation software [online]. Available: oregonstate.edu/dept/eli/june1998.html (5th May 2008).

Hurford J. R. 2004. Language beyond our grasp: What mirror neurons can, and cannot, do for language evolution. In Kimbrough Oller D. and Griebel U. (eds.): 297-313.

IBM 2008. Speech viewer III [online]. Available: www-134.ibm.com/webapp/wcs/stores/servlet/CategoryDisplay?categoryId=2629283&storeId=826&catalogId=-826&langId=826 (5th May 2008).

International Phonetic Association 1999. *Handbook of the International Phonetic Association: A Guide to the Use of the Phonetic Alphabet*. Cambridge University Press.

IPA – The International Phonetic Association 2007. Vowels [online]. Available: www.arts.gla.ac.uk/ipa/ipa.html (5th May 2008).

KayPENTAX 2008. KayPENTAX [online]. Available: www.kayelemetrics.com (5th May 2008).

Keller E. (ed.) [1994] 1999. *Fundamentals of Speech Synthesis and Speech Recognition*. Chippenham: John Wiley & Sons.

Kimbrough Oller D. & Griebel U. (eds.) 2004. *Evolution of Commmunication Systems: A Comparative Approach.* Cambridge: The MIT Press.

Korhonen P. (ed.) 2002. *The Phonetics Symposium 2002*. Espoo: Teknillinen Korkeakoulu.

Kuhl P. 1991. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics.* 50(9) 93-107.

Kungliga Tekniska Högskolan 2005. The 3D vocal tract project [online]. Available: www.speech.kth.se/multimodal/vocaltract.html (5th May 2008).

Leeson R. 1975. *Fluency and Language Teaching*. London: Longman Group Limited.

Lenneberg E. H. 1967. *Biological Foundations of Language*. New York: John Wiley & Sons, Inc.

Liberman A.M. & Mattingly I.G. 1985. The motor theory of speech product revised. *Cognition*. 21: 1-36.

Micro Video Corporation 2008. Video Voice: Speech training system [online]. Available: www.videovoice.com (5th May 2008).

Mishra S. & Sharma R.C. (eds.) 2004. *Interactive Multimedia in Education and Training*. Idea Group Publishing.

Määttä T. 1983. *Hur Finskspråkiga Uppfattar Svenskans Vokaler*. Kungälv: Umeå Universitetet.

Nehaniv C. & Dautenhahn K. (eds) 2002. *Imitation in Animals and Artifacts*. Cambridge: The MIT Press.

Nummenmaa L. 2004. *Käyttäytymistieteiden Tilastolliset Menetelmät*. Helsinki: Kustannusosakeyhtiö Tammi.

O’Shaughnessy D. 1999. *Speech Communications Human Machine 2e*. New York: John Wiley & Sons.

Paganus A., Mikkonen V.-P., Mäntylä T., Nuuttila S., Isoaho J., Aaltonen O. & Salakoski T. 2006. The Vowel Game: Continuous real-time visualization for pronunciation learning with vowel charts. In Salakoski T., Ginter F., Pyysalo S. & Pahikkala T. (eds.): 696-703.

Paganus A. 2007. *Reaaliaikainen Visuaalinen Palaute Vokaalien Ääntämisen Harjoittelussa – Vokaalipeli*. Turku: Turun Yliopisto.

Pedersen U., Rosenberg-Wolff C. & Uddström I. 1996a. “*Va Va De Du Sa?” Gröna Boken.* Helsinki: Helsinki University Press.

Pedersen U., Rosenberg-Wolff C. & Uddström I. 1996b. “*Va Va De Du Sa?” Blåa Boken.* Helsinki: Helsinki University Press.

Perkell J.S. & Nelson W.L. 1987. Articulatory targets and speech motor control: A study of vowel production. In Grillner S., Lindblom B., Lubker J. & Persson A. (eds.): 187-203.

Peterson G.E. & Barney H.L. 1952. Control methods in a study of vowels. *The Journal of the Acoustical Society of America*. 24(2): 175-184.

PROTRAIN 2005. Protrain [online]. Available: www.protrainsys.com (5th May 2008).

Rabiner L. R. & Schafer R. W. 1978. *Digital Processing of Speech Signals*. Prentice-Hall.

Raimo I., Savela J. & Aaltonen O. 2002. The Turku vowel test. In Korhonen P. (ed.): 45-52.

Rizzolatti G. & Craighero L. 2004. The Mirror-neuron system. *Annu. Rev. Neurosci*. 27: 169-192.

Rosner B. S & Pickering J. B. 1994. *Vowel Perception and Production*. Midsomer Norton: Oxford University Press.

Rossing T. D, Moore R. F & Wheeler P.A. [1983] 2001. *The Science of Sound*. Addison-Wesley Publishing Company.

Salakoski T., Ginter F., Pyysalo S. & Pahikkala T. (eds.) 2006. *Advances in Natural Language Processing*. Berlin: Fintal.

Skinner B. F. 1957. *Verbal Behavior*. New York: Appleton-Century-Crofts.

SPSS Inc. 2008. SPSS | Statistical analysis [online]. Available: www.spss.com (5th May 2008).

Tiger DRS, Inc. 1999. Dr. Speech software [online]. Available: www.drspeech.com (5th May 2008).

Zhang F. 2004. Using an interactive feedback tool to enhance pronunciation in language learning. In Mishra S. and Sharma R.C. (eds.): 377-399.

Wiik K. 1980. *Puheen Artikulaation ja Akustiikan suhde 1*. Turku: Turun Yliopiston Suomalaisen ja yleisen kielitieteen laitoksen opintomonisteita.

Wiik K. 1998. *Fonetiikan Perusteet*. Helsinki: WSOY.

Öster A-M. 2006. *Computer-Based Speech Therapy Using Visual Feedback with Focus on Children with Profound Hearing Impairments*. Stockholm: KTH Computer Science and Communication.

Appendix A

The table below shows the 10 treatment group members' 10 average hits of the vowel /ʉ/ in the start test, the 4th exercise and in the end test.

1.		START		1.	4th	Exercise		1.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	312	1632,25	663,25	1	351,818	715,545	259,455	1	386,625	980,125	242,25
2	364,333	1947,33	979,667	2	366,273	786,455	238	2	388,556	815,556	183
3	399,8	1043,4	218,2	3	378	756,333	232,778	3	411,636	815	206,455
4	281	1937	968	4	390,5	868,4	140	4	391,692	857,692	149,077
5	300,375	1878,38	909,375	5	382,077	812	178,154	5	391,818	815	184,636
6	500	1250	344	6	377,8	884,2	274,7	6	365,4	796,5	194,7
7	314,667	1879,75	911,083	7	401,8	958	133,8	7	399,286	867,929	164,714
8	281	1875	906	8	382,5	835,375	160,125	8	383,364	831,909	166,818
9	390,5	1445	484,5	9	379,615	792,846	197	9	391,909	758,273	234,273
10				10	366,833	788,75	200,833	10	380,364	743,909	242,182
Aveg:	349,297	1654,23	709,342	Aveg:	377,722	819,79	201,484	Aveg:	389,065	828,189	196,81
StDev:	72,0229	334,754	292,989	StDev:	13,7806	69,6113	49,0699	StDev:	12,0691	65,8712	33,6177
CoV:	0,20619	0,20236	0,41304	CoV:	0,03648	0,08491	0,24354	CoV:	0,03102	0,07954	0,17081
2.		START		2.	4th	Exercise		2.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	267,222	1169,78	275,111	1	287,889	1107,33	177,111	1	295,091	1192,91	224,636
2	287,2	1143,4	214,6	2	299,7	924,7	67,3	2	281	1263	294,786
3	287,2	1293,5	325,4	3	277,556	909,444	64,8889	3	288,154	1261,77	293
4	296,5	1280,9	312	4	281	954,556	47,3333	4	287,2	1077,7	109,8
5	284,444	1440,56	471,556	5	293,917	1031	119,917	5	299,083	1090,75	123,333
6	287,889	1496,11	527,333	6	287,889	1221,89	264	6	286,636	1124,73	157,818
7	276,571	1437,29	468,571	7	281	933,125	41,75	7	283,583	1077,83	110,5
8	288,75	1366,88	397,875	8	286,636	1053,45	169,455	8	295,308	1098,08	130
9	271,7	1405,9	438	9	275,364	1079,18	145,091	9	290,3	1162,2	193,8
10	284,875	1562,13	593,125	10	284,444	930,111	60	10	283,818	1190	221,636
Aveg:	283,235	1359,64	402,357	Aveg:	285,539	1014,48	115,685	Aveg:	289,017	1153,9	185,931
StDev:	8,80339	136,534	119,26	StDev:	7,39853	102,219	72,8048	StDev:	5,86876	71,4766	71,0797
CoV:	0,03108	0,10042	0,2964	CoV:	0,02591	0,10076	0,62934	CoV:	0,02031	0,06194	0,38229
3.		START		3.	4th	Exercise		3.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	299,6	1018,5	69,5	1	288	878,111	110,556	1	275,833	942,333	57
2	281	827,5	142,5	2	277,9	977,6	166,8	2	267,714	901,286	84
3	281	843	127	3	283,818	931,545	180,273	3	277,125	988	193
4	281	978,667	24,6667	4	265,5	909,75	72,625	4	281	886,25	85
5	312	906	64	5	281	871,4	101,5	5	286,636	860,545	110,273
6	281	757,75	211,25	6	265,5	941	250,75	6	296,5	843,5	127,5
7	406	1374,5	447,5	7	277,556	819	151,444	7	287,2	1052,8	110,6
8	281	750	219	8	293,4	968,5	158,2	8	281	1087,2	120,8
9	281	893,2	81,6	9	290,3	877,8	93,2	9	288,75	1034,63	93,75
10	257,75	781	192,5	10	281	973,5	45	10	302,7	1099,5	130,7
Aveg:	296,135	913,012	157,952	Aveg:	280,397	914,821	133,035	Aveg:	284,446	969,604	111,262
StDev:	41,0715	185,344	121,583	StDev:	9,40174	52,7487	60,2648	StDev:	10,2246	95,6524	36,6651
CoV:	0,13869	0,203	0,76975	CoV:	0,03353	0,05766	0,453	CoV:	0,03595	0,09865	0,32954

4.		START		4.	4th	Exercise		4.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	305,8	943,2	42,6	1	316,846	968,385	92,1538	1	302,7	927,6	152,636
2	324,6	987,2	122,667	2	304,846	927,385	56	2	289,857	995,286	91,7143
3	349,6	943,2	62,6	3	316,429	1035,29	77,4286	3	306,364	877,455	94,3636
4	357,545	948,455	69	4	288,154	1023,69	116	4	312	990,786	47,3571
5	352,6	1021,6	79,9	5	287,2	974,667	78,2667	5	301,667	991,75	69,6667
6	395,583	1095,92	160,917	6	316,133	991,333	67,0667	6	286,636	957,182	69,8182
7	372,538	1052,54	113,385	7	288,75	868,938	107,188	7	292,273	925,636	53,6364
8	380	999,5	94	8	299,167	1015,08	63,5	8	286,813	1114,81	151,125
9	368,909	968,273	72,7273	9	292,625	890,125	97,5	9	297,909	1036,36	104,091
10	397,545	1016,73	111,545	10	295,091	971,091	74,4545	10	309,615	1011,62	66,2308
Aveg:	360,472	997,661	92,9341	Aveg:	300,524	966,598	82,9558	Aveg:	298,583	982,849	90,0639
StDev:	29,1927	50,4778	34,7629	StDev:	12,2231	55,7298	19,6006	StDev:	9,34765	66,2446	37,1282
CoV:	0,08098	0,0506	0,37406	CoV:	0,04067	0,05766	0,23628	CoV:	0,03131	0,0674	0,41224
5.		START		5.	4th	Exercise		5.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	281	1062	94	1	274,111	1048,33	89,6667	1	281	1025,5	87
2	327,5	1140,5	173,5	2	294,286	1084,29	118	2	289,455	1170,09	203,545
3	327,5	1062,5	97	3	307,571	1021,86	76,1429	3	300,727	1045,09	85,4545
4	312	1015,5	58,5	4	287,2	1065,3	105,7	4	312,2	1137	170,2
5	312	1046,5	78	5	299,083	1072,5	105,25	5	292,071	1068,79	102,857
6	281	952,5	35,5	6	270,667	1017	64,3333	6	283,214	1073,29	136,2
7	288,75	1085,5	120	7	316	1058,13	110,25	7	301,667	1098,5	129,833
8	296,5	1062,17	122,167	8	275,833	1015,33	72,5833	8	288,75	1046,58	84,1667
9	250	999,5	64,5	9	292,273	1059,18	98,0909	9	296,5	1034,2	67,5
10				10	312,143	1022,14	62	10	315,875	1030,75	90,125
Aveg:	297,361	1047,41	93,6852	Aveg:	292,917	1046,41	90,2017	Aveg:	296,146	1072,98	115,688
StDev:	25,2454	53,7323	41,2409	StDev:	16,0928	25,3828	20,2112	StDev:	11,6011	48,5504	43,6936
CoV:	0,0849	0,0513	0,44021	CoV:	0,05494	0,02426	0,22407	CoV:	0,03917	0,04525	0,37768
6.		START		6.	4th	Exercise		6.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	419,286	1115,71	282,429	1	412,778	1485,78	533,889	1	478,833	1598,5	654,167
2	281	1343	374	2	455,9	1593,3	645,3	2	468,3	1540,2	595,2
3	437	1031	150	3	464,5	1558,13	614,75	3	437,25	1507,25	557
4	426,667	989	211,333	4	468,333	1593,33	649,167	4	481	1481	543
5	447,667	1140	290,429	5	452,833	1523	576,75	5	456	1406	476,4
6	418,2	1168,4	257,6	6	448,545	1474,09	531,909	6	448,143	1510,64	564,357
7	459,429	1294,14	365,286	7	482,222	1541,22	600,667	7	443,4	1468,4	519
8	405,8	987	165	8	440	1488,27	540	8	459,286	1423,86	483
9	426,667	1104	195,333	9	475,538	1490,08	552,308	9	426,667	1379,83	434,333
10	458	1187,17	277,75	10	456,625	1503,5	560,875	10	369,5	1468,5	511,667
Aveg:	417,971	1135,94	256,916	Aveg:	455,728	1525,07	580,561	Aveg:	446,838	1478,42	533,812
StDev:	51,2371	119,17	77,036	StDev:	19,6727	44,5466	44,5494	StDev:	32,2899	65,1937	63,5021
CoV:	0,12259	0,10491	0,29985	CoV:	0,04317	0,02921	0,07673	CoV:	0,07226	0,0441	0,11896
7.		START		7.	4th	Exercise		7.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	437,25	1257,5	410,25	1	364,333	1135	214,333	1	265,5	1046,5	86,5
2	281	1031	64	2	362,4	1168,4	217,2	2	366,75	1265	312
3	343,5	1296,5	401,25	3	343,5	1156	194,5	3	402	1140,38	213,625
4	343	1062	102	4	331	1087,2	134,6	4	388	1093,43	165,714
5	375	1000	81	5	320	1304,25	343,5	5	382,625	1077,75	148,5
6	390,5	1093,5	155,5	6	387,2	1112,2	172	6	398,125	1140	200,125
7	296,5	1202,5	239,5	7	281	1328	359	7	333	1093,33	140,667
8	406	1031	530	8	328	1265,25	309,75	8	380	1098,5	230,333
9				9	359	1124,5	179,5	9	374,667	1062	142
10				10				10	395,5	1140,17	201,167
Aveg:	359,094	1121,75	247,938	Aveg:	341,826	1186,76	236,043	Aveg:	368,617	1115,71	184,063
StDev:	53,5414	114,111	177,693	StDev:	31,1668	88,8984	80,8081	StDev:	41,2999	61,9077	62,0982
CoV:	0,1491	0,10173	0,71669	CoV:	0,09118	0,07491	0,34235	CoV:	0,11204	0,05549	0,33737

8.		START		8.	4th	Exercise		8.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	296,5	1027,9	68,1	1	305,429	1033,29	74,7143	1	278,615	1023,62	100,214
2	312	1004	67,5714	2	295,308	1088,62	134,154	2	320,467	1018,33	114,875
3	288,75	903,333	79,25	3	309,692	1011,62	60,6923	3	307,231	1095,69	133,846
4	298,222	961,222	32,6667	4	295,308	1033,38	69,8462	4	301,118	1012,59	105,882
5	323,364	1022,27	98,9091	5	305,8	1076,67	109,8	5	292,071	1051	84,3571
6	309,692	1009,23	59,0769	6	301,667	1103,73	138,667	6	300,929	1044,29	89,3571
7	312	1057,31	91,5385	7	307,867	1072,53	105,667	7	300,818	991,091	56,7273
8	307,231	1026,08	61,6154	8	305,111	1185,33	219,667	8	302,9	1212	270,909
9	312	1024,7	65,2	9	301,667	1039,2	74,4	9	303,143	1030,86	66,4286
10	314,385	1102,85	135,615	10	307,933	1095,53	130,867	10	307,571	1157,93	189,643
Aveg:	307,414	1013,89	75,9543	Aveg:	303,578	1073,99	111,847	Aveg:	301,486	1063,74	121,224
StDev:	10,1283	53,1731	27,7928	StDev:	5,05042	49,8619	47,619	StDev:	10,7891	70,7591	64,5499
CoV:	0,03295	0,05244	0,36592	CoV:	0,01664	0,04643	0,42575	CoV:	0,03579	0,06652	0,53248
9.		START		9.	4th	Exercise		9.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	292,273	883,273	103,455	1	273,25	1239,17	281,167	1	271,7	1090,3	144,8
2	301,667	924,25	62,5	2	289,263	1124,68	170,158	2	259,538	922,692	202,154
3	294,357	932,714	72,7143	3	261,923	1098,08	139,692	3	275,833	913,667	198,667
4	293,4	915,1	58,8	4	273,706	1141,29	174,941	4	267,438	989,75	116,313
5	292,923	1002,08	60,5385	5	275,529	1179,82	212,176	5	273,846	992,385	47,1538
6	357,941	1069,53	164,176	6	268,235	1117,24	165,412	6	274,8	997,6	46,8
7	312	968,385	33,7692	7	265,5	1048,79	137,143	7	272,733	939,133	58,2667
8	288,75	921,5	63,25	8	274,474	1162,47	196,263	8	276,867	1039,2	77,4667
9	290,3	896,7	82,2	9	267,714	1131,36	171,786	9	276,867	1101,6	156,063
10	307,571	1035,5	73,5714	10	268,944	1098,67	135,889	10	269,375	1011,31	67,6875
Aveg:	303,118	954,903	77,4974	Aveg:	271,854	1134,16	178,463	Aveg:	271,9	999,764	111,537
StDev:	20,7362	62,0866	35,3156	StDev:	7,50202	51,8532	43,8696	StDev:	5,34203	64,5017	60,7107
CoV:	0,06841	0,06502	0,4557	CoV:	0,0276	0,04572	0,24582	CoV:	0,01965	0,06452	0,54431
10.		START		10.	4th	Exercise		10.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	374,857	1298,86	338,571	1	355,231	1129,46	171,154	1	334,357	1142,57	181,571
2	361,286	1258,57	299,714	2	332,917	1096	134,5	2	343,313	1138,25	180,313
3	330,6	1037	350,333	3	347,714	1084,5	130,143	3	327,643	1153,57	192,786
4	331,625	1222,25	407,556	4	340,769	1059,69	109,308	4	340,364	1175,91	213,182
5	335,5	1026,88	81,5	5	387,75	1080,25	148,833	5	346,4	1112	185,1
6	360,889	1228,89	270,333	6	365,3	1090,2	153,3	6	343,273	1155,82	195,364
7	362,4	1174,6	224,6	7	330,8	1005,8	70,4	7	320,636	1087,64	125
8	430,222	1232,22	298,667	8	338,857	1122,29	167,5	8	298,714	1093,29	128,857
9	364,333	1254,83	306,333	9	336,571	1037,57	115,5	9	340,75	1075,08	139,333
10	421,5	1148	226,75	10	352,143	1050,93	109,714	10	335,667	1085,67	195,538
Aveg:	367,321	1188,21	280,436	Aveg:	348,805	1075,67	131,035	Aveg:	333,112	1121,98	173,704
StDev:	34,472	92,5433	89,1172	StDev:	17,4236	37,9922	30,9284	StDev:	14,4207	35,5286	31,0351
CoV:	0,09385	0,07788	0,31778	CoV:	0,04995	0,03532	0,23603	CoV:	0,04329	0,03167	0,17867

Appendix B

The table below shows the 10 treatment group members' 10 average hits of the word /hʉs/ in the vowel diagram in the start test, the 4th exercise and in the end test.

1.		START		1.	4th	Exercise		1.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	374,75	1270,5	319,583	1	380,364	1099,18	185,727	1	421,417	1085,5	228,75
2	351,375	1214,63	262,75	2	392,222	982,333	184,222	2	340,111	1093,33	308,889
3	383,182	1113,27	286	3	376,929	1030,79	193,429	3	421,5	1105,13	295,222
4	380,083	1137,58	283,5	4	378,111	1048,33	197,667	4	413,75	987,875	190,875
5	356,583	1181,92	278,583	5	383,273	1005,45	121,545	5	385,222	1225,33	292,333
6	372,385	1153,46	224,462	6	379,286	1062,14	197,429	6	395,556	1260,22	193,25
7	381,923	1105,38	183,462	7	374,857	1102,29	185,286	7	405,889	992,556	135,333
8	384,3	1115,5	180,4	8	395,667	921,5	109	8	399	1201	262
9	380,545	1198,55	261,727	9	420,091	1116,27	201,091	9	423,111	1096,78	211,556
10	388,143	1053,14	182,143	10	405,875	1042,63	142	10	378,333	971,778	118
Aveg:	375,327	1154,39	246,261	Aveg:	388,667	1041,09	171,74	Aveg:	398,389	1101,95	223,621
StDev:	12,1739	63,207	50,2944	StDev:	14,7994	60,0002	34,2202	StDev:	25,708	101,32	66,4235
CoV:	0,03244	0,05475	0,20423	CoV:	0,03808	0,05763	0,19926	CoV:	0,06453	0,09195	0,29704
2.		START		1.	4th	Exercise		1.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	284,1	1305,9	336,9	1	274,8	940,2	64,2	1	285,769	1081,38	142,692
2	296,5	1692,5	724,167	2	287,889	1204,44	238,889	2	283,818	1073,55	135,727
3	285,429	1575,43	606,429	3	281	1034	125,1	3	291,333	1038,75	104
4	276,571	1374,71	406,857	4	281	892,429	134,143	4	306,545	1065,09	102,455
5	277,125	1159,88	192,125	5	287,889	996,111	84,6667	5	281	1100,33	136,889
6	284,444	1228,78	259,889	6	292,364	928,636	138,636	6	277,9	1131	166
7	287,2	1199,4	235,4	7	290,3	1084,2	224,2	7	288	1037,78	104
8	285,429	1155,71	190	8	296,5	956,75	96,625	8	274,111	1037,67	118,111
9	281	1156,13	188,5	9	293,4	1015,3	130,3	9	288,75	987,75	77,5
10	284,1	1146,5	250,4	10	283,818	1047,82	154,727	10	288,75	968,375	119,875
Aveg:	284,19	1299,49	339,067	Aveg:	286,896	1009,99	139,149	Aveg:	286,598	1052,17	120,725
StDev:	5,59641	193,057	187,304	StDev:	6,69415	90,4877	55,8961	StDev:	8,8373	49,1939	25,3595
CoV:	0,01969	0,14856	0,55241	CoV:	0,02333	0,08959	0,4017	CoV:	0,03084	0,04675	0,21006
3.		START		3.	4th	Exercise		3.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	322,333	1041,33	76,5	1	322,333	1051,83	89,6667	1	299,6	1268,2	300,6
2	296,5	1109	140,875	2	307,571	1106,86	142,571	2	324,4	1305,8	337,6
3	319,75	999,75	84,25	3	332,667	1083	118,667	3	349,4	1143,6	181
4	312	1187	219,818	4	340,2	1184,1	220,6	4	250	1687	719
5	301,667	1166,33	201,5	5	327,5	1152	186	5	312	1005,8	44,8
6	332,667	1051,67	102,5	6	331,375	1230	264,625	6	343,4	1124,6	168,2
7	318,2	1105,8	139,8	7	318,889	1162,78	195,778	7	318,889	1100,33	137,778
8	299,6	1006	52,2	8	312,083	1189,75	225,083	8	329,714	1191,57	224,857
9	304,25	1179,25	214,75	9	308,125	1163,75	196,125	9	327,5	1129,83	164,167
10	322,333	1093,33	126,333	10	320,857	1133,43	168,143	10	298,714	1084,43	125,857
Aveg:	312,93	1093,95	135,853	Aveg:	322,16	1145,75	180,726	Aveg:	315,362	1204,12	240,386
StDev:	11,9785	68,885	59,6394	StDev:	10,9191	53,3127	52,5413	StDev:	28,2867	190,82	188,218
CoV:	0,03828	0,06297	0,439	CoV:	0,03389	0,04653	0,29072	CoV:	0,0897	0,15847	0,78298

4.		START		4.	4th	Exercise		4.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	312,333	1395,33	430,667	1	316	1155,75	198	1	351	1163,63	208,5
2	399,8	1199,6	251,6	2	327,667	1051,67	94,1667	2	314,818	1150,18	218
3	368,6	1205,8	246,4	3	324,4	1199,8	232,6	3	331,818	1002,55	80,4545
4	349,6	1218,4	258	4	316	1160	212	4	343,25	1049,08	164,071
5	378,75	1230	273,625	5	338	1072,5	124,333	5	328,769	1095,69	140,846
6	378,75	1202,75	250,75	6	332,778	1162,89	206	6	319	1065,67	107
7	369,182	1198,64	243	7	312	1079,27	141,091	7	324,5	1096,5	130,4
8	343,286	1133,43	172,857	8	323,875	1194,88	252,875	8	330,7	1130,9	166,6
9	352,429	1209,43	252,857	9	334	1071,4	144,6	9	359	1390,5	427,5
10	356,571	1182,71	223,143	10	348,909	1070,64	149,273	10	320,455	1090,55	124,182
Aveg:	360,93	1217,61	260,29	Aveg:	327,363	1121,88	175,494	Aveg:	332,331	1123,52	176,755
StDev:	23,9571	67,6228	65,7853	StDev:	11,3825	57,7732	51,7213	StDev:	14,4583	105,286	97,8591
CoV:	0,06638	0,05554	0,25274	CoV:	0,03477	0,0515	0,29472	CoV:	0,04351	0,09371	0,55364
5.		START		5.	4th	Exercise		5.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	270,667	999,667	50,5	1	288,75	1148	180	1	286,167	1109	145
2	287,4	1012,2	139,833	2	257,75	1077,5	117	2	301,667	1156	188,667
3	287,2	1043,4	84,2	3	281	1125	157	3	276,571	1124,57	157,714
4	270,667	978,667	57,3333	4	296,5	1171,5	202,75	4	325,571	1053,14	90,7143
5	265,5	1062	109,75	5	288,75	1116,75	148,5	5	300,818	1141,82	182
6	260,333	947,333	48,6667	6	296,5	1156	187,5	6	315,444	1114,22	148,667
7	250	1031	79	7	281	1228,67	263,333	7	316,429	1142,71	175,143
8	406	1250	300	8	281	1145,33	180,333	8	288,75	1156	187,75
9	250	1031	79	9	286,167	1182	217,333	9	287,2	1162,3	196,4
10				10	312	1218	249	10	281	1176,5	210,5
Aveg:	283,085	1039,47	105,365	Aveg:	286,942	1156,88	190,275	Aveg:	297,962	1133,63	168,255
StDev:	48,0471	86,2435	78,629	StDev:	14,0469	45,795	44,8549	StDev:	16,7196	35,518	34,4125
CoV:	0,16973	0,08297	0,74625	CoV:	0,04895	0,03959	0,23574	CoV:	0,05611	0,03133	0,20453
6.		START		6.	4th	Exercise		6.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	452,8	1221,5	315,7	1	430,8	1496,6	548,6	1	451,111	1461,33	524,667
2	390,25	1265	438,6	2	409,625	1550,5	597,5	2	416,444	1489,33	539,667
3	428,143	1173,71	345,5	3	425,727	1462,55	514	3	432,714	1423,86	482,143
4	437,125	1249,63	347	4	422,909	1502,36	552,545	4	414,364	1530,91	576,364
5	421,333	1369,5	439,833	5	428,143	1490,71	539,714	5	425,5	1640,25	689,25
6	418,2	1330,8	399,6	6	426,667	1546,42	599,667	6	418,4	1499,5	552,9
7	405,857	1401,29	467,429	7	444,111	1541,22	597,556	7	437	1508	562
8	390,25	1495,63	552,25	8	374,714	1580	622,857	8	390,3	1565,2	608
9	405,857	1379,14	439,286	9	402,7	1490,3	534,8	9	413,625	1530,88	579,375
10	436,75	1226,25	306	10	427,7	1430,8	486,5	10	410,143	1464	514,571
Aveg:	418,657	1311,24	405,12	Aveg:	419,31	1509,15	559,374	Aveg:	420,96	1511,33	562,894
StDev:	20,7448	100,343	77,3207	StDev:	19,3283	45,162	43,5731	StDev:	16,6882	60,8403	57,1652
CoV:	0,04955	0,07652	0,19086	CoV:	0,0461	0,02993	0,0779	CoV:	0,03964	0,04026	0,10156
7.		START		7.	4th	Exercise		7.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	281	1406	437,25	1	359	1171,5	214,5	1	318,4	1193,2	233,6
2	367	1343,5	382,75	2	302	1260,33	295,333	2	330,143	1343,29	390,429
3	312,333	1354	411,667	3	335,75	1366,75	402,25	3	318,4	1406	440,4
4	333	1280,67	320,667	4	368,4	1374,6	423,6	4	281	1239	274,333
5	301,667	1291,33	323	5	338,333	1395,67	434,833	5	388,143	1222,86	273,286
6	318,6	1387,2	424,4	6	385	1364	409,333	6	379,833	1291,33	346,167
7	328	1359	392,5	7	321,5	1327,6	368,3	7	333	1270,33	314,833
8	343	1187	222	8	351,25	1280,75	324	8	343,5	1338	387,5
9	312,333	1260	312	9	260,333	1374,67	407,667	9	312,5	1249,5	287,5
10	327,833	1291,17	333,333	10	317,5	1317,17	355	10	337,2	1174,6	222,2
Aveg:	322,477	1315,99	355,957	Aveg:	333,907	1323,3	363,482	Aveg:	334,212	1272,81	317,025
StDev:	23,4465	66,2229	65,9713	StDev:	35,9134	69,0621	68,9256	StDev:	31,4573	72,6254	72,2566
CoV:	0,07271	0,05032	0,18534	CoV:	0,10756	0,05219	0,18963	CoV:	0,09412	0,05706	0,22792

8.		START		8.	4th	Exercise		8.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	322,444	1183,56	218,111	1	281	1033,73	142,545	1	275,833	1020,5	69,1667
2	336,9	1112,2	186	2	287,889	1016,89	78,3333	2	296,5	1108,75	143,25
3	317,167	1077,83	110,333	3	293,4	1052,8	95,2	3	320,857	1102,29	142,714
4	329,222	996,111	66,2222	4	292,273	1167,18	199,091	4	304,25	1108,75	177,111
5	332,667	1037,89	82,6667	5	304,923	1038,15	109,692	5	312	1143,33	177,133
6	308,9	1193,4	228,3	6	293,917	1088,17	122,333	6	293,4	1121,4	153,1
7	307,571	1053,29	92,5714	7	291,333	1101,17	139,667	7	326,182	1172,91	205,818
8	329,222	1142,11	176,444	8	325,571	1140,21	182,571	8	306,364	1090,45	123,636
9	371,6	1199,7	245,7	9	296,5	1101,25	134,917	9	306,455	1047,91	100,818
10	360,778	1270,44	311,444	10	296,5	1233,93	266,786	10	324,154	1076,46	145,5
Aveg:	331,647	1126,65	171,779	Aveg:	296,331	1097,35	147,114	Aveg:	306,599	1099,28	143,825
StDev:	20,7543	86,1775	81,4239	StDev:	11,9771	67,8989	55,7231	StDev:	15,4648	44,075	39,4152
CoV:	0,06258	0,07649	0,474	CoV:	0,04042	0,06188	0,37878	CoV:	0,05044	0,04009	0,27405
9.		START		9.	4th	Exercise		9.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	314,385	980,308	142,308	1	272,545	1121,82	155,818	1	273,25	1025,58	127,923
2	365,4	1077,7	131,2	2	274,8	1149,73	183,4	2	267,222	1024	176,333
3	346,364	1016,64	101	3	296,643	1182,57	217,643	3	271,941	1067,59	127,765
4	383,091	1065	138,364	4	283,385	1115,08	147,385	4	265,5	1071,4	114,1
5	333,769	1141,54	183,615	5	283,818	1175,73	207	5	275,833	1051,83	89,1667
6	345,615	1175,08	213,308	6	302,5	1151,88	187,938	6	291,333	1067,42	102
7	340,455	1130,36	173,364	7	321,3	1190,2	223,1	7	274,222	1166,33	205,333
8	319,833	1090,83	127,833	8	257,154	1083,85	125,846	8	278,182	1124,64	158,727
9	332	1167	210,143	9	272,143	1122,36	156,643	9	276,867	1212,13	250
10	323,455	1144,45	187,091	10	271,313	1177,5	214,688	10	301,071	1220,64	255,143
Aveg:	340,437	1098,89	160,823	Aveg:	283,56	1147,07	181,946	Aveg:	277,542	1103,16	160,649
StDev:	21,1127	64,8236	37,9264	StDev:	18,6488	35,2249	33,9448	StDev:	10,8522	73,4847	59,6543
CoV:	0,06202	0,05899	0,23583	CoV:	0,06577	0,03071	0,18657	CoV:	0,0391	0,06661	0,37133
10.		START		10.	4th	Exercise		10.		END	
/hUs/	F1	F2	Eucl. Dist.	/hUs/	F1	F2	Eucl. Dist.	/hUs/	F1	F2	Eucl. Dist.
1	307,714	1396,86	430	1	372,467	1270,53	316,8	1	374,545	1241,09	283,455
2	320	1292,63	327,375	2	365,571	1200,57	243,071	2	335,5	1249,67	284,5
3	358,9	1321,3	365,5	3	433,375	1335,56	392,438	3	351,632	1284,21	327,105
4	339,375	1374,63	408,125	4	378,167	1289,61	334,944	4	365,353	1266,18	307,765
5	330,6	1337	369,6	5	376,857	1265,21	309,714	5	372,357	1227,29	271,643
6	350,333	1284,33	324	6	367,857	1216,07	264,357	6	386,438	1187,13	244,125
7	362	1193,6	235,4	7	368,4	1243,47	284,667	7	333,769	1203,92	239,538
8	327,7	1337,2	371,2	8	386,722	1303,39	349	8	350,278	1289,56	327,5
9	353,889	1336,44	376,444	9	394,95	1345,05	391,85	9	336,5	1276,43	311,214
10	378,857	1218,29	265,571	10	354,353	1290,18	330,765	10	349,313	1276,94	320,75
Aveg:	342,937	1309,23	347,322	Aveg:	379,872	1275,96	321,761	Aveg:	355,568	1250,24	291,76
StDev:	21,7228	64,0586	60,488	StDev:	21,9478	47,0667	49,371	StDev:	18,3024	35,1061	32,5307
CoV:	0,06334	0,04893	0,17416	CoV:	0,05778	0,03689	0,15344	CoV:	0,05147	0,02808	0,1115

Appendix C

The table below shows the 10 control group members' 10 average hits of the vowel /ʉ/ in the vowel diagram in the start test, the 4th exercise and in the end test. The control group was never exposed to The Vowel Game's real-time visual feedback.

11.		START		11.	4th	Exercise		11.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	312	1593	624	1	334,933	1305,87	340,933	1	332,667	1287,89	321,333
2	343	1625	657	2	312,25	1585,5	619	2	338,714	1267,57	302
3	312	1554,25	585,75	3	337,2	1309	345,4	3	366,133	1260	305,067
4	312	1687	718,5	4	343,333	1531	564,667	4	352,143	1584,43	617,286
5	330,6	1555,8	587,4	5	351	1609	641,625	5	343,167	1148,17	185,75
6	281	1593	624	6	347,643	1332,29	367,286	6	345,357	1218,5	258,5
7	336,8	1662,2	694	7	349	1374,55	409,545	7	332,917	1504,75	537,583
8	312	1760	791,667	8	336,8	1381	413,4	8	336,429	1426	459
9	301,667	1666,33	697,333	9	348,333	1413,58	447	9	332,667	1398,08	430,75
10				10	321,538	1201,46	234,692	10	321,143	1455	488
Aveg:	315,674	1632,95	664,406	Aveg:	338,203	1404,32	438,355	Aveg:	340,134	1355,04	390,527
StDev:	18,935	67,4853	67,5632	StDev:	12,7511	132,247	131,696	StDev:	12,4574	139,262	136,685
CoV:	0,05998	0,04133	0,10169	CoV:	0,0377	0,09417	0,30043	CoV:	0,03662	0,10277	0,35
12.		START		12.	4th	Exercise		12.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	260,333	976,167	73,6667	1	252,583	1119,58	158	1	252,818	1107,55	166,273
2	291,333	1062	94	2	257,75	1218,25	253,25	2	250	1124,5	223,5
3	260,333	1037,89	101,333	3	256,2	1224,4	260,8	3	250	1131,44	200,444
4	281	1018,4	59,6	4	260,333	1156	193	4	263,286	1044,29	87,5714
5	288,75	1155,75	394,5	5	281	1263,44	317	5	258,857	1013	217,571
6	270,667	1254,83	291	6	250	1135,33	173	6	260,333	899	89,3333
7	281,167	911,167	123	7	255,167	1015,33	70,5	7	281	937	37
8	265,5	1171,33	336,667	8	265,5	1226,25	261,5	8	250	950,429	59,2857
9	281	918,4	61,2	9	250	885	154,333	9	250	1083	124,333
10	265,5	953	87	10				10	255,167	1098,5	138,5
Aveg:	274,558	1045,89	162,197	Aveg:	258,726	1138,18	204,598	Aveg:	257,146	1038,87	134,381
StDev:	11,5153	115,68	126,975	StDev:	9,71225	121,014	75,1755	StDev:	9,6994	84,711	66,4204
CoV:	0,04194	0,1106	0,78285	CoV:	0,03754	0,10632	0,36743	CoV:	0,03772	0,08154	0,49427
13.		START		13.	4th	Exercise		13.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	423,111	930,222	150,444	1	382,5	874,75	138,25	1	437	937	140
2	423,111	1024	156,444	2	395,333	989,333	103,333	2	444,75	1007,5	154,75
3	400,5	926,667	132,167	3	390,167	1010	177	3	281	968	19
4	374,571	839	155	4	426,667	963	128,5	4	362	1168,4	241
5	394	968,375	130,875	5	399,6	1005,8	126	5	359	1093,5	157,5
6	405,75	929,5	145	6	405,875	929,25	125,125	6	363	1042,5	136,5
7	383,429	946	102,571	7	354	1187,33	283,333	7	380,455	1050,73	190,091
8	312,333	926,667	65	8	340,364	897,364	94,6364	8	374,5	937	104
9	312	999,5	37,5	9	369,091	965,545	120,182	9	382,5	913,75	152,5
10	402,333	930	136,444	10	395,5	942,5	152,167	10	382,5	1132,5	204,25
Aveg:	383,114	941,993	121,145	Aveg:	385,91	976,488	144,853	Aveg:	376,67	1025,09	149,959
StDev:	40,3351	49,7327	40,502	StDev:	25,4585	86,2184	53,9276	StDev:	45,0045	87,6863	60,07
CoV:	0,10528	0,0528	0,33433	CoV:	0,06597	0,08829	0,37229	CoV:	0,11948	0,08554	0,40058

14.		START		14.	4th	Exercise		14.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	250	937	66,25	1	274,8	984	191,1	1	275,833	1090,67	359,333
2	250	1031	79	2	264,308	1004,23	272	2	290,3	890,4	129
3	265,5	796,5	177	3	266,909	1081,91	138,818	3	288,75	1025,67	319,25
4	265,5	1192,33	351,333	4	263,778	1173,22	274,667	4	276,231	925,154	185,538
5	265,5	866,75	108,5	5	272,286	928,286	316,286	5	258,455	928,636	218
6	260,333	885	96,3333	6	265,5	1176,67	399,333	6	271,462	1331,31	444,154
7	263,286	1289,71	494,714	7	264,091	1218,27	454,909	7	260,333	978,667	329,5
8	281	828	144	8	283,818	1013,82	296	8	275,833	895,5	97,8333
9	273,25	898,25	78,75	9	289,455	1127,45	365,818	9	268,083	947,583	299,25
10	274,8	1143,6	244	10	283,818	775,182	194,182	10	308,9	1040,2	146,6
Aveg:	264,917	986,815	183,988	Aveg:	272,876	1048,3	290,311	Aveg:	277,418	1005,38	252,846
StDev:	9,97497	169,007	141,086	StDev:	9,67093	135,25	98,8829	StDev:	15,1512	131,993	113,909
CoV:	0,03765	0,17126	0,76682	CoV:	0,03544	0,12902	0,34061	CoV:	0,05461	0,13129	0,45051
15.		START		15.	4th	Exercise		15.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	333,692	764,231	209,077	1	340,923	994,846	63,2308	1	319,813	976,313	54,5
2	326,091	1641,55	673	2	374,769	1098,23	153	2	314,385	1088,46	120,846
3	331,154	1593,38	625,154	3	382,5	1028,42	102,167	3	336,643	1160,29	201,357
4	330,667	839,2	138,667	4	374,818	971,182	82,8182	4	381,133	1097,27	152,533
5	316,429	1843,21	874,286	5	359,167	1109	152,5	5	396,385	1143,92	210,462
6	334,8	1357,87	390,133	6	337,917	1062,08	116,083	6	334	1019,82	82,4706
7	325,286	1347,79	379,786	7	372,308	997,308	88,5385	7	353,125	991,813	72,9375
8	337,067	997,533	114,533	8	420,091	1062,09	157,364	8	363,316	1073,58	137,737
9	338,385	1057,38	97,8462	9	383,455	906	117,818	9	346,765	913,176	77,7647
10	322,333	929,25	54,75	10	406	1116,27	182,273	10	368	999,643	91
Aveg:	329,59	1237,14	355,723	Aveg:	375,195	1034,54	121,579	Aveg:	351,356	1046,43	120,161
StDev:	6,97198	372,844	284,371	StDev:	25,6385	67,7699	38,4815	StDev:	26,3273	78,9579	54,4815
CoV:	0,02115	0,30138	0,79942	CoV:	0,06833	0,06551	0,31651	CoV:	0,07493	0,07545	0,45341
16.		START		16.	4th	Exercise		16.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	302,9	927,8	198,3	1	292,273	1073,45	112,545	1	299,083	997,167	36,5833
2	314,818	937,182	162,636	2	289,455	1036,55	73,4545	2	290,3	1049,6	81,7
3	281	808,9	161,5	3	284,875	1058,25	102,625	3	303,545	1076,09	108,909
4	281	812	157,727	4	284,444	850,333	120	4	291,333	1020,44	65,4444
5	308,9	1015,4	49	5	293,4	1121,6	162,6	5	281	1093,33	127,556
6	317,636	982,545	52,6364	6	286,167	1231,33	263,167	6	281	1052,7	90,3
7	301,667	921,5	53	7	286,727	1224,09	259,818	7	291,333	1004,83	47,8333
8	304,25	1033,5	71,25	8	321	1080	116,286	8	273,846	1045,23	87,0769
9	299,167	882,417	107,333	9	284,444	1048,22	83	9	283,385	1040,46	78,4615
10	296,5	815,2	154,1	10	274,357	1024,21	76,7143	10	274,8	943,3	44,5
Aveg:	300,784	913,644	116,748	Aveg:	289,714	1074,8	137,021	Aveg:	286,963	1032,32	76,8364
StDev:	12,3283	83,1681	56,5256	StDev:	12,1806	107,702	70,5091	StDev:	9,86084	43,0017	28,9644
CoV:	0,04099	0,09103	0,48417	CoV:	0,04204	0,10021	0,51459	CoV:	0,03436	0,04166	0,37696
17.		START		17.	4th	Exercise		17.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	250	978,667	52,6667	1	273,25	968,5	46,75	1	291,333	1072,33	108
2	257,75	984	299,2	2	265,5	1056,83	96,8333	2	260,333	1291,33	372,333
3	268,6	1180,6	214,4	3	281	1249,5	282	3	317,5	1181,83	327,833
4	265,5	1213,17	246,667	4	262,4	1174,4	230	4	268,6	887,2	88,2
5	257,75	1132,25	168,25	5	250	1072,67	188,333	5	281	1322,67	356
6	270,667	1176,67	209,667	6	274,8	1024,8	89,8	6	296,5	953	79
7	250	1187	223	7	260,333	1749,33	781	7	273,25	866,75	139
8	273,25	1234,25	266,25	8	265,5	976,25	66,75	8	312	1374,5	470
9	265,5	1101,25	204,2	9	260,333	1301,67	337,667	9	270,667	1156	377,333
10	270,667	1135,33	169	10	268,6	1005,8	93,2	10	287,889	978,889	53,2222
Aveg:	262,968	1132,32	205,33	Aveg:	266,172	1157,98	221,233	Aveg:	285,907	1108,45	237,092
StDev:	8,56085	88,6917	67,2037	StDev:	8,78857	237,188	219,899	StDev:	18,8404	184,809	156,997
CoV:	0,03255	0,07833	0,3273	CoV:	0,03302	0,20483	0,99397	CoV:	0,0659	0,16673	0,66218

18.		START		18.	4th	Exercise		18.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	359	1433,25	608,889	1	406	1218	270	1	265,5	1499,5	531,5
2	250	1234	291,5	2	250	2031	1063	2	390,5	1234	287
3	406	1552	592,333	3	343,667	978,667	184	3	250	1968	1000
4	368,6	1599,6	637,8	4	500	1312	397	4	374,5	1280,5	337
5	359,125	1640,13	684,5	5	468,5	1374,5	440,5	5	390,333	1213	269,667
6	382,5	1664	703,25	6	398	1155,75	253	6	419,286	1392,43	453,143
7	325,571	1624,57	657,286	7	351,25	1577,75	253,667	7	390,25	1554,5	601
8	359	1624,5	662,25	8	447,667	1447,67	502,667	8	447,333	1333	393,667
9	250	1531	564	9	390,5	1546,5	608,5	9	491,75	1390,25	467,25
10				10	412	1274,6	337	10	412	1399,6	448
Aveg:	339,977	1544,78	600,201	Aveg:	396,758	1391,64	430,933	Aveg:	383,145	1426,48	478,823
StDev:	55,3129	136,151	123,992	StDev:	70,8547	288,139	257,382	StDev:	74,4259	218,792	210,549
CoV:	0,1627	0,08814	0,20658	CoV:	0,17858	0,20705	0,59727	CoV:	0,19425	0,15338	0,43972
19.		START		19.	4th	Exercise		19.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	312,273	948,455	139,636	1	416,417	1231,33	287,5	1	426,667	1319,92	373,667
2	343,125	1042,38	89,125	2	431,833	1330,33	385,083	2	452,5	1166,33	252,417
3	365,2	905,9	107,1	3	454,222	1273,78	341,889	3	437,071	1251,93	314,786
4	424,6	849,8	173,2	4	434,786	1222,79	287,643	4	406	1206,23	261,077
5	489,333	947,333	199,333	5	441,769	1213,69	283	5	394	1131,85	197,692
6	421,5	1031	137,25	6	437	1180,43	251,929	6	400,273	1067,82	160,455
7	401,429	1022,14	174,143	7	453,769	1256,92	326,692	7	450,563	1118,75	215,625
8	413,667	1182,08	282,75	8	478	1292,92	370,385	8	422,4	1130,73	212,6
9	359,25	925,375	124,25	9	425	1317,08	374,154	9	456,231	1213,38	290,769
10	374,462	1345,69	391,154	10	432,231	1158,31	238,462	10	462,375	1196,75	285,25
Aveg:	390,484	1020,02	181,794	Aveg:	440,503	1247,76	314,674	Aveg:	430,808	1180,37	256,434
StDev:	50,3165	146,936	91,714	StDev:	17,578	56,7775	52,3893	StDev:	24,7954	73,1067	62,756
CoV:	0,12886	0,14405	0,50449	CoV:	0,0399	0,0455	0,16649	CoV:	0,05756	0,06194	0,24473
20.		START		20.	4th	Exercise		20.		END	
/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/U/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	285,769	1187,08	220,846	1	300,727	1110,36	142,273	1	327,5	1132,5	165,5
2	302,462	982,692	61,6923	2	287,889	1065,56	102,444	2	330,6	1206	238,8
3	300,929	1042,07	126,143	3	339,125	1265,38	298,375	3	335,333	1176,67	210,75
4	301,667	926,75	71,25	4	334,545	1178,45	211,909	4	343,273	1124,45	164,727
5	307,231	975,615	64	5	315,444	1232,33	263,667	5	360,778	1214,89	253,444
6	314,286	1053,21	93,3571	6	333,7	1171,6	205,3	6	346,2	1156	192,1
7	309,786	1271,93	303,286	7	346,2	1208,9	244,1	7	345,75	1197,5	233,917
8	328,846	1141,54	175,385	8	336,111	1270,56	303,333	8	355,9	1243,2	281,5
9	312	990,643	67,3571	9	315,1	1174,6	206,3	9	360,667	1218,33	256,222
10	298,714	906	70	10	312	1234,08	265,25	10	352,846	1215,92	254,615
Aveg:	306,169	1047,75	125,332	Aveg:	322,084	1191,18	224,295	Aveg:	345,885	1188,55	225,158
StDev:	11,3605	118,302	82,6492	StDev:	18,794	65,6478	64,7798	StDev:	11,9209	39,6538	40,2512
CoV:	0,03711	0,11291	0,65944	CoV:	0,05835	0,05511	0,28882	CoV:	0,03446	0,03336	0,17877

Appendix D

The table below shows the 10 control group members' 10 average hits of the /hʉs/-word in the vowel diagram in the start test, the 4th exercise and in the end test. The control group was never exposed to The Vowel Game's real-time visual feedback.

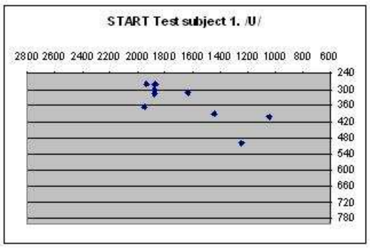
11.		START		11.	4th	Exercise		11.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	365,857	1267,43	332	1	325,571	1267,29	305,143	1	356,929	1231,86	271,357
2	355,25	1327,75	379,875	2	339,5	1386,25	421,875	2	324,6	1449,6	482,6
3	392,571	1187,29	271,857	3	371,6	1259	300,4	3	312	1507,25	539,25
4	359,1	1140,3	219,4	4	340	1374,67	409,556	4	351,375	1323,88	365,125
5	377,7	1196,6	299,2	5	366,182	1508,27	447	5	368,6	1356	394,9
6	374,667	1162,67	244,333	6	374,667	1343,27	385,4	6	353,917	1392,92	428,583
7	437,222	1225,33	295,778	7	348,75	1411,08	447,25	7	349,6	1284	320,2
8	374,571	1258,43	322,857	8	380	1155,83	216	8	359,1	1431	466,4
9	381	1380,8	422,8	9	389	1414,27	455,182	9	361,667	1452,75	487,333
10	412,2	1180,8	258,6	10	354,909	1400,27	435,909	10	365,077	1547,77	582,231
Aveg:	383,014	1232,74	304,67	Aveg:	359,018	1352,02	382,371	Aveg:	350,286	1397,7	433,798
StDev:	25,1485	76,2977	62,2788	StDev:	20,4866	100,106	81,1209	StDev:	18,0921	99,442	97,1887
CoV:	0,06566	0,06189	0,20441	CoV:	0,05706	0,07404	0,21215	CoV:	0,05165	0,07115	0,22404
12.		START		12.	4th	Exercise		12.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	281	968,444	101	1	281	972,714	39,1429	1	284,444	1062,11	148,889
2	292,625	882,375	97,625	2	304,25	1105	150,25	2	288,75	1168,83	207,25
3	329,333	1076,11	133,556	3	305,222	1204,56	246	3	309,182	1067,73	101,455
4	316,571	1013	139,571	4	290,3	984	75,7	4	304,25	960,5	57,625
5	281	995,625	70,625	5	312	1010	51,3333	5	289,455	1008,18	112,727
6	317,25	1106,42	144,75	6	287,2	1112	145	6	291,333	1030,83	68,5
7	270,667	958	62	7	312	1099	138,091	7	265,5	1115,2	157
8	296,5	952,75	40	8	339,25	1061,88	118,5	8	321,3	1102,7	138,3
9	281	1039,57	95,8571	9	304,25	960,5	74	9	271,7	1090,2	131,8
10	308,125	1022,88	155,375	10	291,333	994,333	51,1667	10	270,667	1034,33	77,5556
Aveg:	297,407	1001,52	104,036	Aveg:	302,681	1050,4	108,918	Aveg:	289,658	1064,06	120,11
StDev:	19,5547	65,1491	38,7709	StDev:	16,7138	79,1166	63,7592	StDev:	17,9287	59,2744	45,9848
CoV:	0,06575	0,06505	0,37267	CoV:	0,05522	0,07532	0,58538	CoV:	0,0619	0,05571	0,38286
13.		START		13.	4th	Exercise		13.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	418,3	1006	219,7	1	399,4	1149,8	267	1	343,25	1390,25	451
2	409,75	980,125	157,5	2	327,75	1312,25	372	2	400,5	1379,83	431,5
3	401,429	950,429	178,143	3	382,5	1343,25	396,25	3	385	1489,33	518,25
4	417,5	917,625	139	4	437	1088,86	220,286	4	452,75	1179,5	272,5
5	432,429	928,143	212	5	406	1281	333,5	5	359	1359	394,5
6	424,6	893,6	147,2	6	327,75	1312,25	345,75	6	405,8	1387,2	457
7	401,286	901,286	142,429	7	392,429	1160,29	256,286	7	426,333	1135	209,333
8	416,333	1202,5	272,667	8	433,375	1171,38	280,875	8	421,5	1531	582,75
9	423,571	964	147,714	9	415,1	1149,6	261,5	9	374,75	1437,25	482,75
10	398,889	947,444	126,667	10	463,083	1057	201,917	10	296,5	1671,5	702,5
Aveg:	414,409	969,115	174,302	Aveg:	398,439	1202,57	293,536	Aveg:	386,538	1395,99	450,208
StDev:	11,2753	89,087	46,4608	StDev:	44,1035	101,365	65,037	StDev:	45,505	156,799	141,425
CoV:	0,02721	0,09193	0,26655	CoV:	0,11069	0,08429	0,22156	CoV:	0,11772	0,11232	0,31413

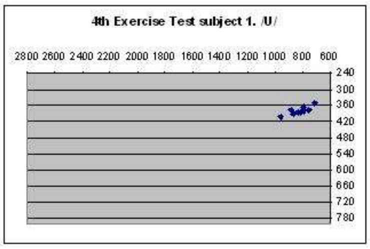
14.		START		14.	4th	Exercise		14.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	289,857	1138	169,571	1	368,909	1059,36	249,545	1	287,889	843,444	142,222
2	312	1234,17	270	2	298,333	912,778	130,222	2	297,909	965,455	71,2727
3	303,143	1097,57	130,286	3	286,167	786,167	183,167	3	294,889	926,556	119,889
4	296,5	1223,5	255,5	4	275,833	978,833	294,833	4	267,222	864,222	120,111
5	304,25	1077,75	109,25	5	294,286	1066,71	245,286	5	260,333	895,5	93
6	312	1214	245	6	334,818	996,727	175	6	285,429	1088,86	220,571
7	312	1187,17	219,833	7	285,429	888,143	135	7	295,273	1087,73	223,636
8	312	1265,25	296,25	8	288,75	917,625	309,125	8	285,429	1026,43	127
9	305,8	1237	269,2	9	285,429	1160,43	264	9	274,8	905,6	164,4
10	305,8	1318,2	349,6	10	307,714	1035,14	84,7143	10	301,667	1114,33	146,667
Aveg:	305,335	1199,26	231,449	Aveg:	302,567	980,192	207,089	Aveg:	285,084	971,812	142,877
StDev:	7,46641	75,2428	75,2497	StDev:	28,512	107,555	76,226	StDev:	13,6826	100,433	49,3854
CoV:	0,02445	0,06274	0,32512	CoV:	0,09423	0,10973	0,36808	CoV:	0,04799	0,10335	0,34565
15.		START		15.	4th	Exercise		15.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	339,25	1199	234,063	1	338,538	1110,15	150,769	1	360,231	1131,77	183,231
2	320,545	1726,82	758,364	2	390,5	1040,4	117,6	2	357,545	1110,36	162,091
3	345,692	1069,31	118,077	3	372,643	1033,21	124,5	3	389,467	1087,13	150,267
4	354,5	1153,57	196,929	4	385,222	1089,89	148,778	4	400,364	1155,91	212,545
5	350,538	1415,46	449,846	5	395,5	1234	281,833	5	382,667	1213,08	258,083
6	341,133	1435,07	468,4	6	388,143	1173,86	224	6	403,333	1202,83	260,417
7	337,364	1587,82	619,636	7	355,867	1147,53	196	7	384,538	1100,54	158,308
8	348,333	1140,33	193,833	8	409,1	1062,1	145,6	8	385,333	1108,92	168,333
9	367,231	1357,69	395,308	9	390,5	1116,88	173,5	9	384,2	1205,8	255,2
10	353,154	1487,62	522	10	370,357	1042,07	133,071	10	401,077	1160,62	227
Aveg:	345,774	1357,27	395,646	Aveg:	379,637	1105,01	169,565	Aveg:	384,876	1147,7	203,547
StDev:	12,4346	214,132	207,98	StDev:	20,6785	65,5863	51,3207	StDev:	15,7201	47,1038	44,4274
CoV:	0,03596	0,15777	0,52567	CoV:	0,05447	0,05935	0,30266	CoV:	0,04084	0,04104	0,21827
16.		START		16.	4th	Exercise		16.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	312	820,857	149	1	304,25	1109,13	140,375	1	296,5	1120,75	154
2	327,5	991,75	122,125	2	337	931	80,8	2	301,667	1062,33	104
3	337	968,4	97,9	3	312	994,667	46,3333	3	307,571	1053,29	117,286
4	334,143	897,143	83,1429	4	312	1430,8	145	4	312	1208	239,667
5	343,125	1003,63	69,75	5	312	963,857	55	5	319,75	1007,5	58,75
6	343	921,75	77,5	6	296,5	1031	69,75	6	303,143	1071	105,857
7	332,667	817,333	156,5	7	295,091	1099	135,727	7	307,571	1164,86	237,286
8	362,75	972,375	90,75	8	355,7	1205,8	260,7	8	300,375	1089,63	126
9	347	976,125	92,625	9	318,2	984	87	9	343	1124,67	161
10	320,857	968,429	64	10	315,3	1012,2	145,7	10	291,333	1093	133,667
Aveg:	336,004	933,779	100,329	Aveg:	315,804	1076,14	116,639	Aveg:	308,291	1099,5	143,751
StDev:	14,2573	68,0841	32,02	StDev:	18,335	148,748	63,6256	StDev:	14,5696	57,7945	57,4549
CoV:	0,04243	0,07291	0,31915	CoV:	0,05806	0,13822	0,54549	CoV:	0,04726	0,05256	0,39968
17.		START		17.	4th	Exercise		17.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	260,333	1041,33	84	1	304,25	1374,75	406	1	281	1062	94,3333
2	260,333	1259,67	293	2	281	1041,33	103	2	294,429	1169,29	208,571
3	265,5	1078	113,5	3	304,25	1077,75	118,75	3	286,167	1228,67	260,333
4	287,2	1374,6	406,6	4	285,429	1280,71	324,714	4	281	1168,6	200,6
5	250	1031	79	5	290	1289,71	327,571	5	385	1208	275
6	302	1499,33	532,667	6	274,8	1118,4	151,2	6	287,4	1193,2	228,6
7				7	299,6	1206	237,8	7	351,25	1210,5	262,25
8				8	296,5	1109	140,5	8	291,333	1249,67	281
9				9	333	1281	313,667	9	288,75	1085,5	117,5
10				10	296,5	1124,5	155,5	10	327,75	1772,75	806,5
Aveg:	270,894	1213,99	251,461	Aveg:	296,533	1190,32	227,87	Aveg:	307,408	1234,82	273,469
StDev:	19,607	195,495	190,625	StDev:	16,1174	111,315	107,888	StDev:	35,5554	198,19	197,886
CoV:	0,07238	0,16104	0,75807	CoV:	0,05435	0,09352	0,47346	CoV:	0,11566	0,1605	0,72361

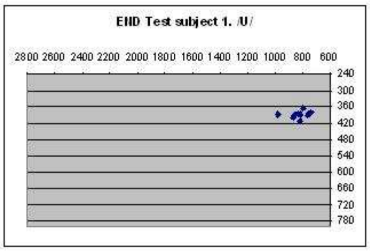
18.		START		18.	4th	Exercise		18.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	398	1695	741,25	1	400,667	1702,83	747,167	1	390,25	1570	612
2	385	1416,33	460,333	2	349,8	1731	765,4	2	419,143	1499,57	566,571
3	374,5	1515,5	557,5	3	437,125	1601,13	655,5	3	398	1601	643,5
4	405,75	1710,5	750,75	4	327,75	1679,25	712,25	4	374,727	1508	547,364
5	312	1500	531	5	412,2	1574,8	621,2	5	374,7	1596,5	633,4
6	312	1312	343	6	401,429	1517,57	560	6	419,143	1664,86	708,714
7	312	1562	594	7	424,4	1462,2	512,8	7	405,6	1631	673,6
8	281	1875	906	8	374,667	1509,83	550	8	364	1687	722,333
9	296,5	1640	671	9	418,2	1562,2	612,8	9	368,6	1774,8	868,333
10				10	365,857	1624,57	663,571	10	374,4	1762,2	799
Aveg:	341,861	1580,7	617,204	Aveg:	391,209	1596,54	640,069	Aveg:	388,856	1629,49	677,482
StDev:	48,2136	169,695	169,871	StDev:	35,3256	88,4927	84,9829	StDev:	20,6495	94,4958	100,449
CoV:	0,14103	0,10735	0,27523	CoV:	0,0903	0,05543	0,13277	CoV:	0,0531	0,05799	0,14827
19.		START		19.	4th	Exercise		19.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	433,556	1551,67	598,222	1	426,583	1351,08	406,333	1	446	1383,57	440
2	468,375	1413,63	489,625	2	436,857	1330	392,357	2	458	1208	286,333
3	400,182	1462,64	511,636	3	432,154	1374,77	430,769	3	430,8	1259	326,9
4	427,6	1562,2	611,3	4	446,3	1377,9	436,9	4	458,9	1259,1	332
5	437,111	1551,78	603,333	5	428,071	1211,64	279,643	5	452,583	1254,92	324,75
6	433,9	1449,7	504,5	6	436,846	1179,85	274,923	6	429,167	1208	277,25
7	474,091	1448,55	512,273	7	410,286	1325,57	380	7	434,333	1233,83	298,75
8	452,75	1494,33	552,833	8	439,462	1389	461,846	8	446,2	1277,8	350
9	491,75	1429,25	500,75	9	443,333	1393,47	460,733	9	450,286	1265,21	339,786
10	452,625	1487,75	546	10	442,636	1317,73	382,455	10	458	1319	402,333
Aveg:	447,194	1485,15	543,047	Aveg:	434,253	1325,1	390,596	Aveg:	446,427	1266,84	337,81
StDev:	26,4074	53,9914	46,5525	StDev:	10,6164	73,5882	66,5615	StDev:	11,3766	52,4372	50,47
CoV:	0,05905	0,03635	0,08572	CoV:	0,02445	0,05553	0,17041	CoV:	0,02548	0,04139	0,1494
20.		START		20.	4th	Exercise		20.		END	
/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)	/hUs/	F1 (Hz)	F2 (Hz)	Eucl. (Hz)
1	296,5	1096,4	174,3	1	330,529	1302,88	341,471	1	312	1363,43	394,571
2	295,182	1119,09	152,636	2	310,188	1267,25	300,188	2	365,714	1432,71	470
3	292,923	1018,85	121,385	3	309,857	1209,5	242,571	3	329	1280,91	313,636
4	312,111	996,111	105,444	4	329,688	1237,81	273,188	4	376,714	1359,07	403,571
5	321,692	1194,23	229,923	5	310	1249,64	284,286	5	355	1357,06	393,5
6	300,154	1115,08	197,077	6	338,167	1253,11	302	6	374,583	1395,67	434,167
7	304,375	1183,19	264,25	7	343,308	1204	260,462	7	327,643	1316,57	350,714
8	325,083	1236,58	269,583	8	349,467	1270,53	306,067	8	334,727	1258,18	292,636
9	294,778	1121,22	152,778	9	330	1321,14	355,929	9	310	1341,27	373,8
10	330,733	1247,47	283,6	10	370	1334,57	373,857	10	322,533	1293,27	327,533
Aveg:	307,353	1132,82	195,098	Aveg:	332,12	1265,04	304,002	Aveg:	340,792	1339,81	375,413
StDev:	14,0722	84,2367	64,0116	StDev:	19,369	43,9451	42,1776	StDev:	25,1981	53,5605	55,2361
CoV:	0,04579	0,07436	0,3281	CoV:	0,05832	0,03474	0,13874	CoV:	0,07394	0,03998	0,14713

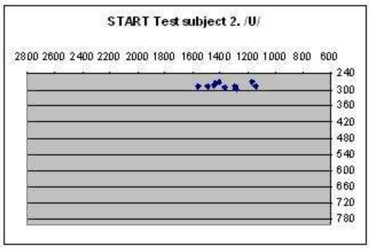
Appendix E

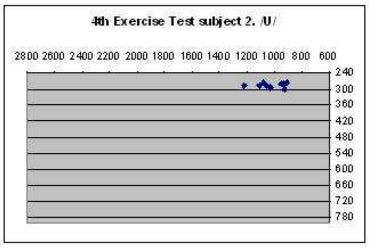
Below appear the visualizations of the 10 treatment group members' 10 average /ʉ/-vowels in the start test, the 4th exercise and the end test. The values are expressed in Hertz.

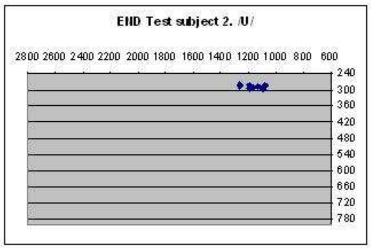


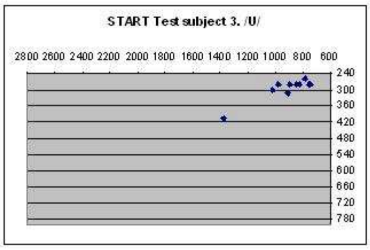


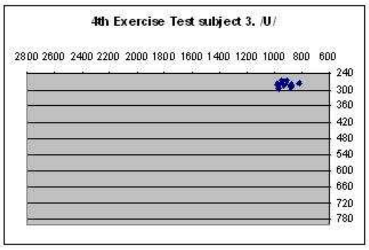


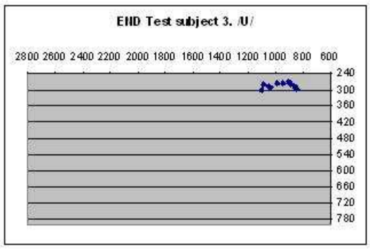


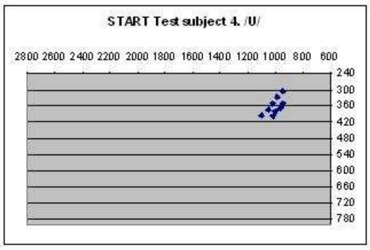


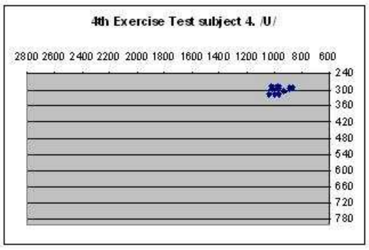


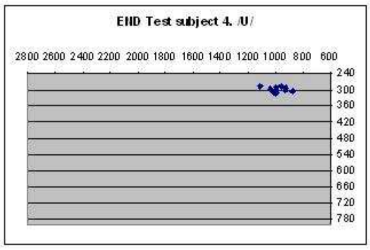


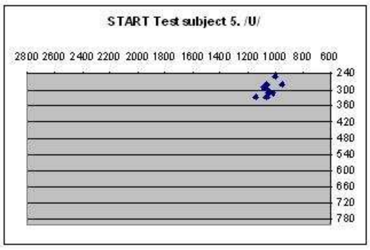


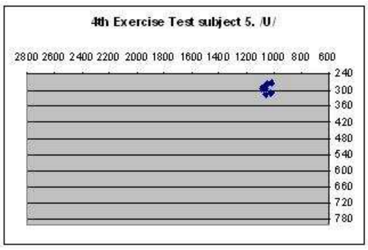


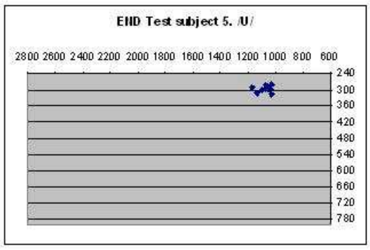


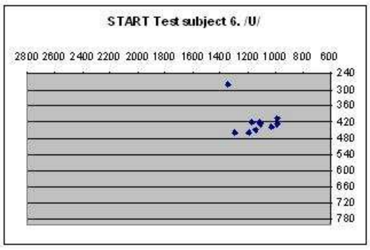


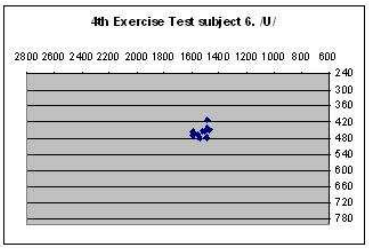


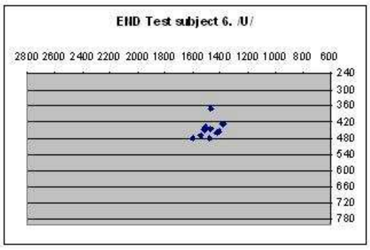


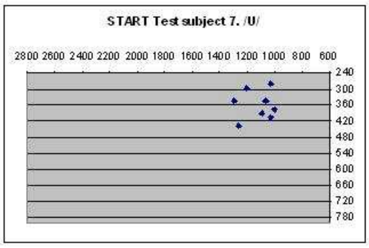


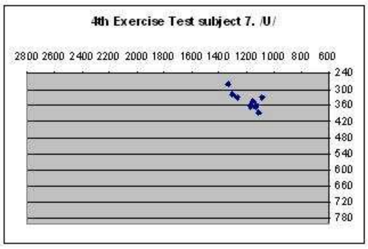


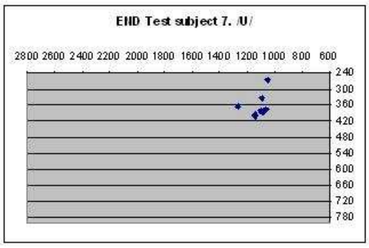

START Test subject 6. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

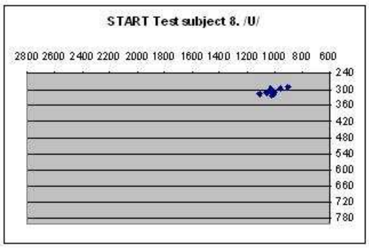

4th Exercise Test subject 6. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

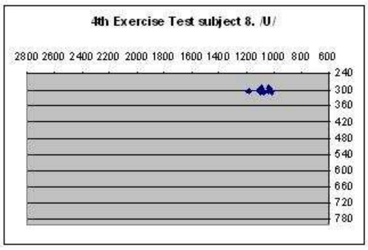

END Test subject 6. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

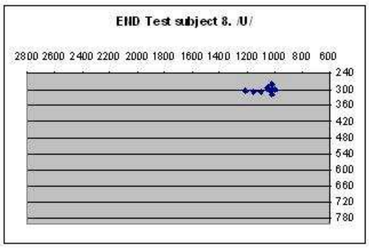

START Test subject 7. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

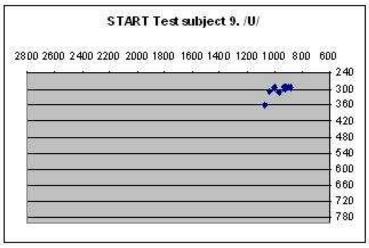

4th Exercise Test subject 7. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

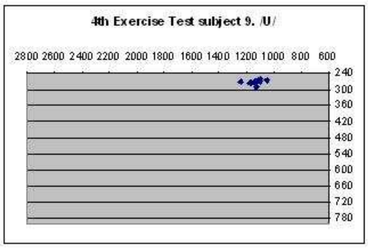

END Test subject 7. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

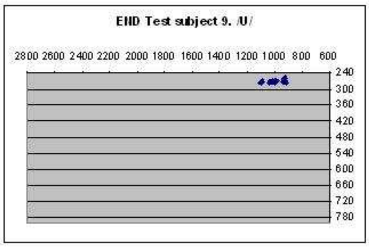

START Test subject 8. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

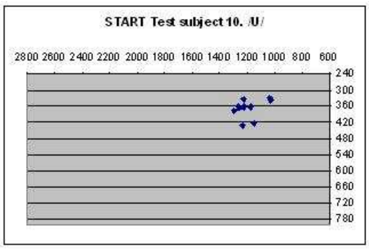

4th Exercise Test subject 8. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

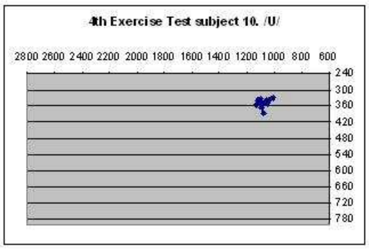

END Test subject 8. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

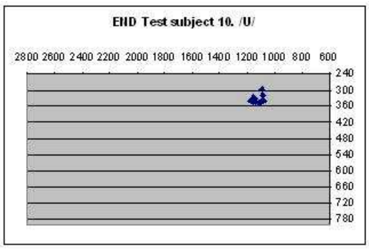

START Test subject 9. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780


4th Exercise Test subject 9. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780


END Test subject 9. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

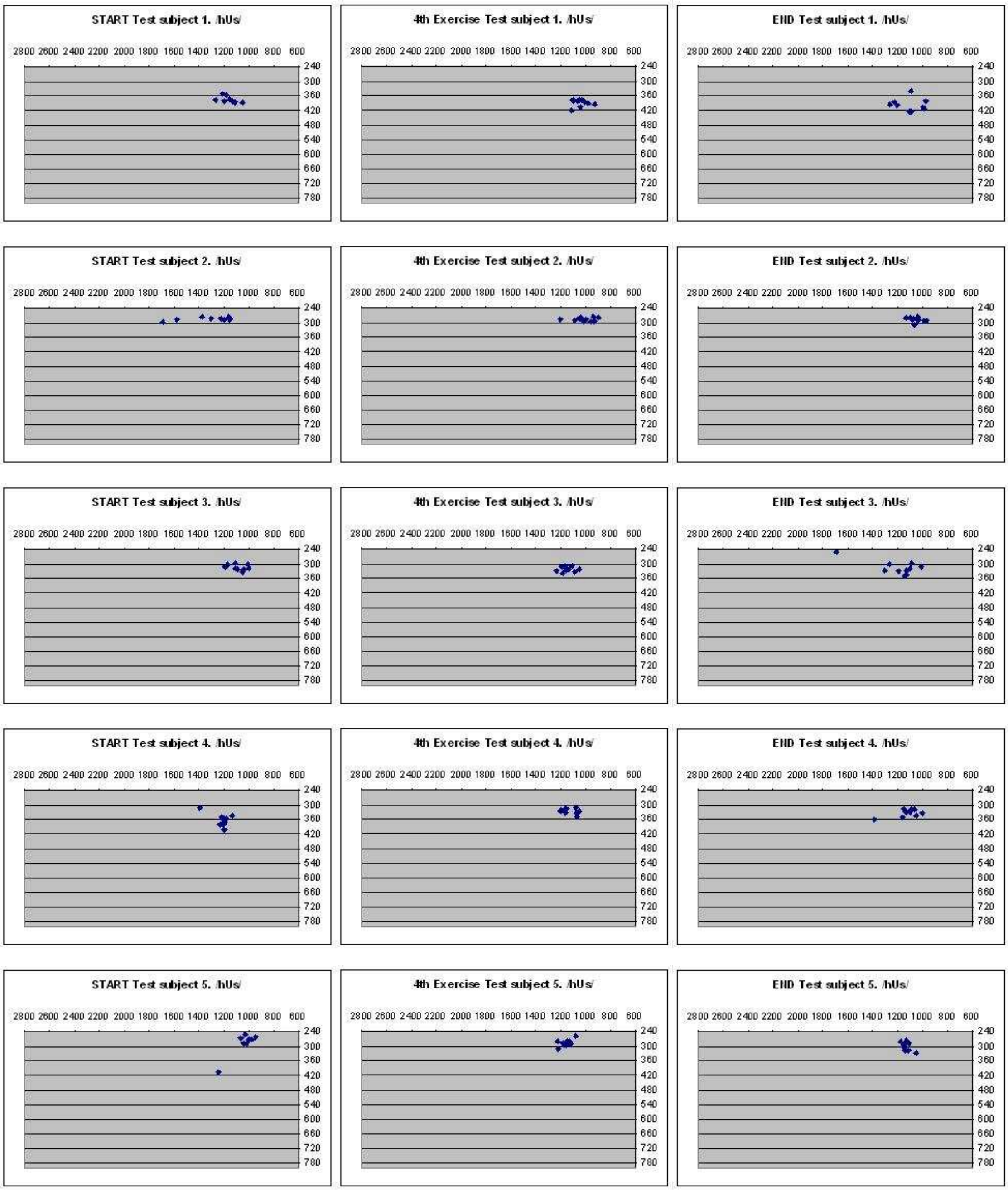

START Test subject 10. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

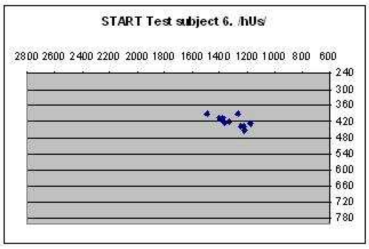
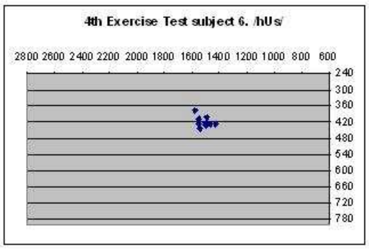
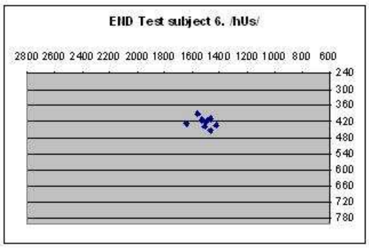
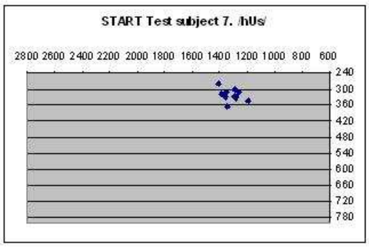
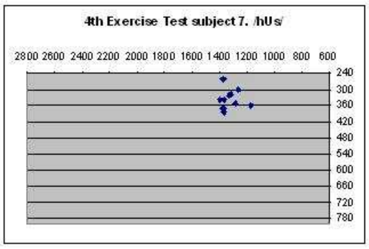
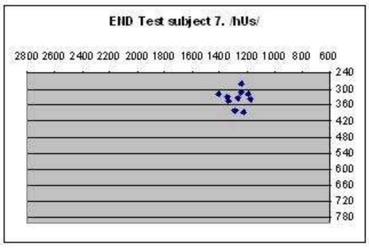
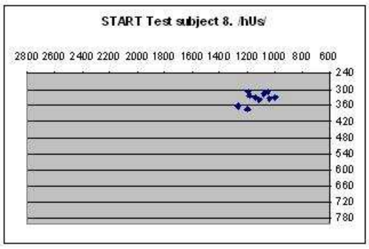
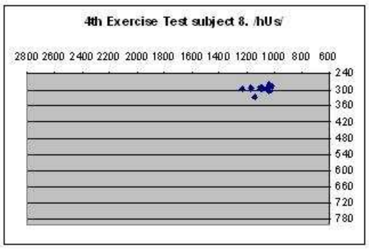
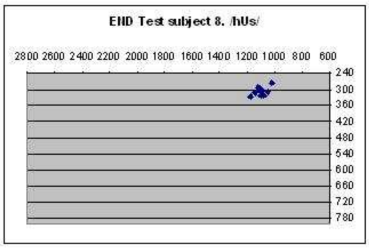
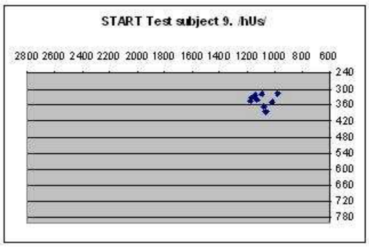
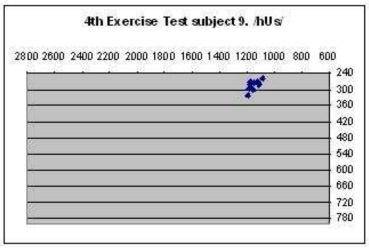
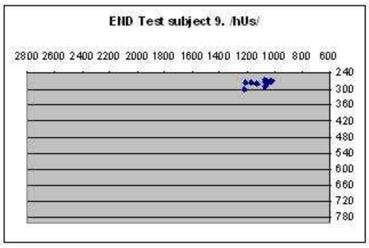
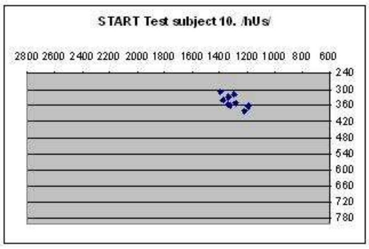
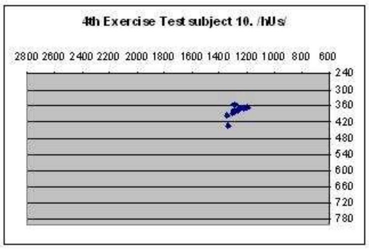
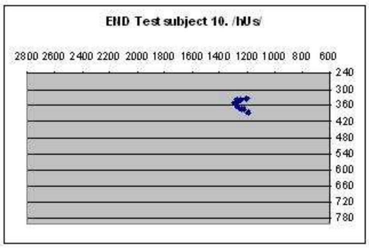

4th Exercise Test subject 10. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780


END Test subject 10. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

Appendix F

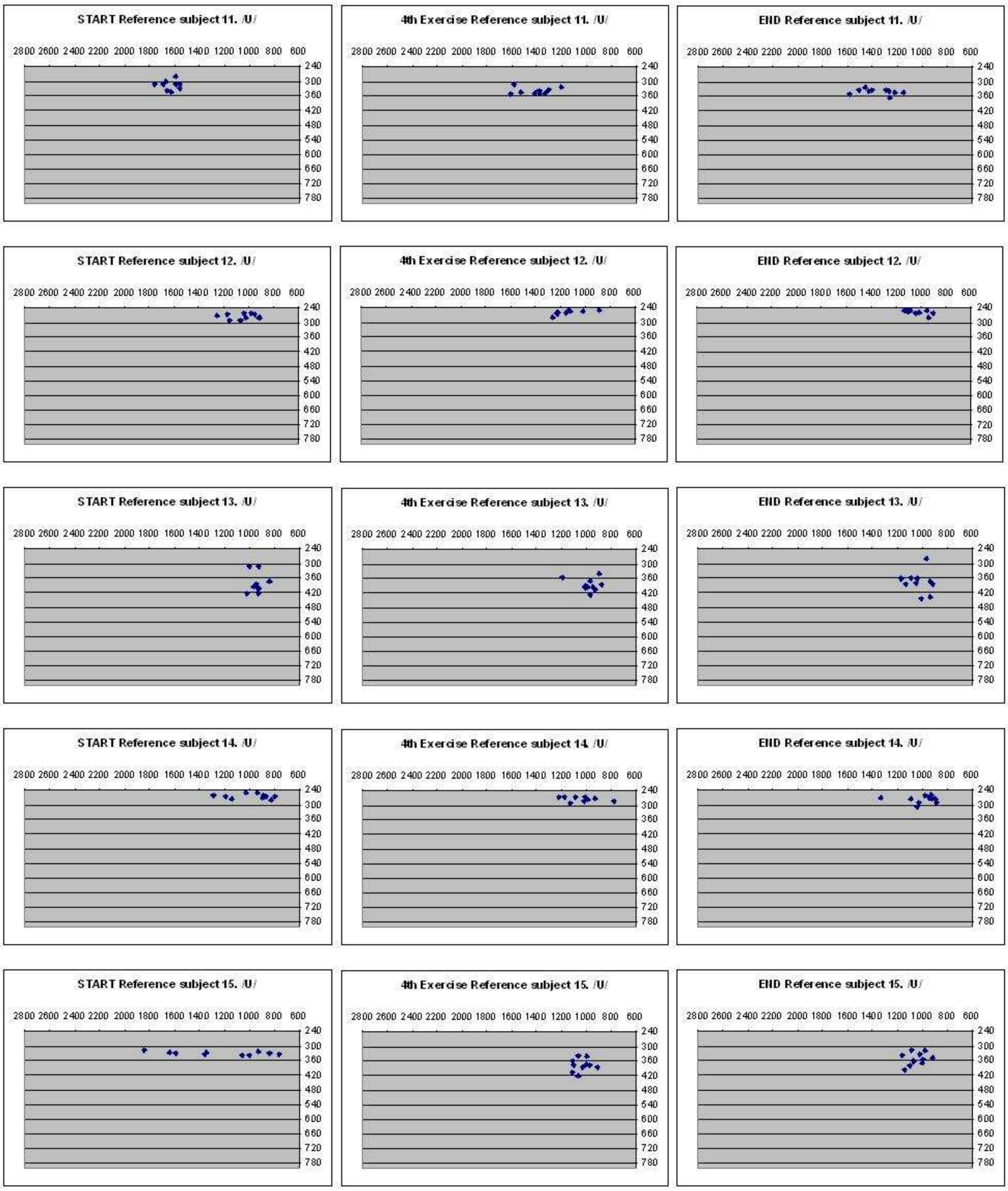
Below appear the visualizations of the 10 treatment group members' 10 average /hʉs/-vowels in the start test, the 4th exercise and in the end test. The values are expressed in Hertz.

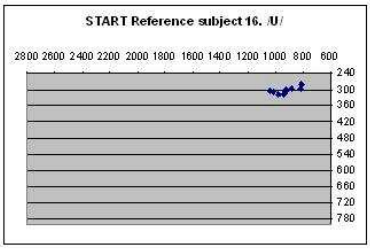


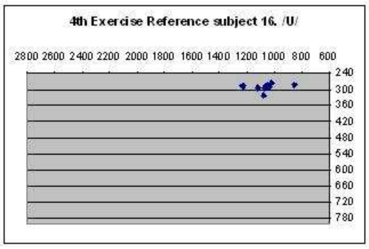

START Test subject 6. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

4th Exercise Test subject 6. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

END Test subject 6. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

START Test subject 7. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

4th Exercise Test subject 7. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

END Test subject 7. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

START Test subject 8. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

4th Exercise Test subject 8. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

END Test subject 8. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

START Test subject 9. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

4th Exercise Test subject 9. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

END Test subject 9. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

START Test subject 10. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

4th Exercise Test subject 10. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

END Test subject 10. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

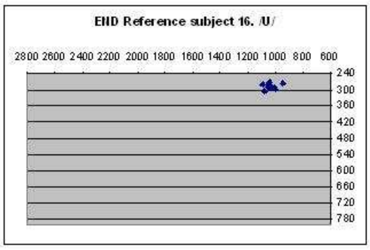
Appendix G

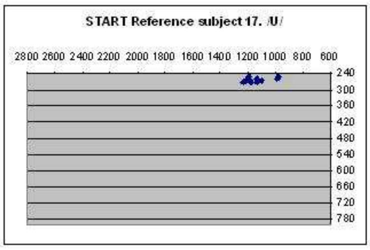
Below appear the visualizations of the 10 control group members' 10 average /ʉ/-vowels in the start test, the 4th exercise and in the end test. The control group was never exposed to The Vowel Game's real-time visual feedback. The values are expressed in Hertz.

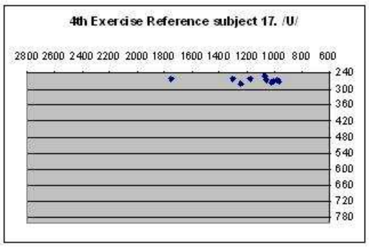


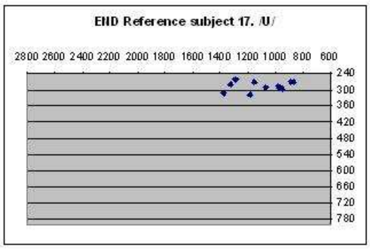

START Reference subject 16. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

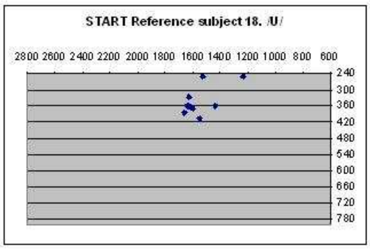

4th Exercise Reference subject 16. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

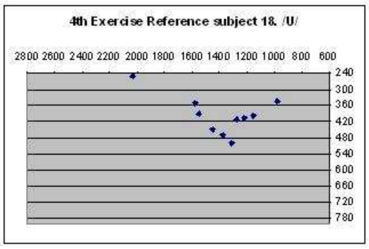

END Reference subject 16. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

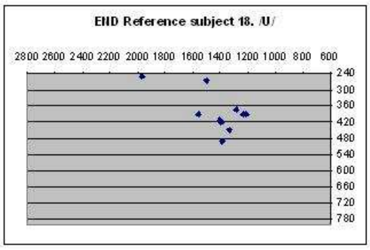

START Reference subject 17. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

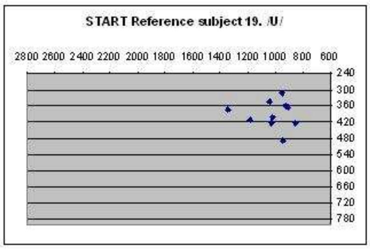

4th Exercise Reference subject 17. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

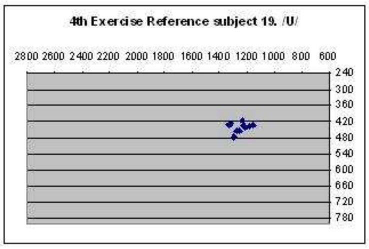

END Reference subject 17. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

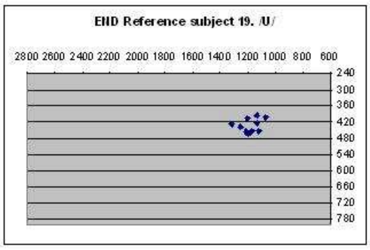

START Reference subject 18. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

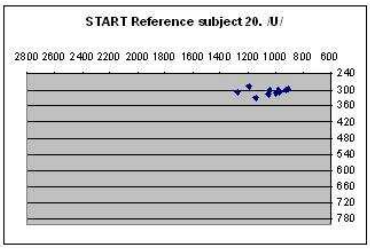

4th Exercise Reference subject 18. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

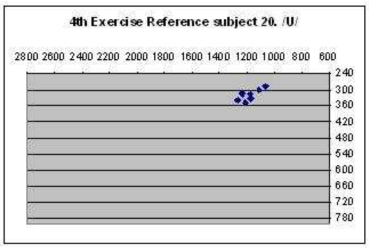

END Reference subject 18. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

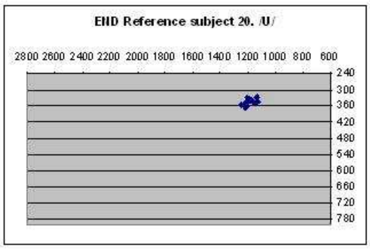

START Reference subject 19. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780


4th Exercise Reference subject 19. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780


END Reference subject 19. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

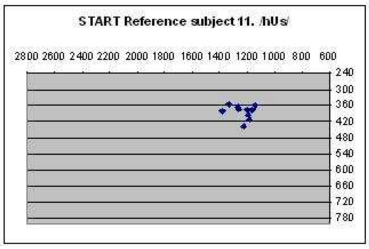

START Reference subject 20. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

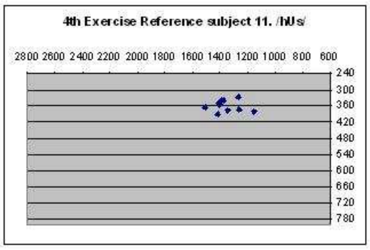

4th Exercise Reference subject 20. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

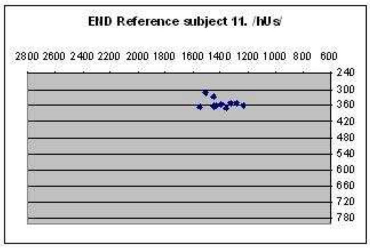

END Reference subject 20. /U/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

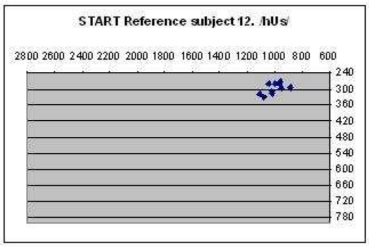
Appendix H

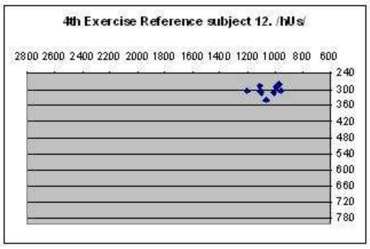
Below appear the visualizations of the 10 control group members' 10 average /hʉs/-vowels in the start test, the 4th exercise and in the end test. The control group was never exposed to The Vowel Game's real-time visual feedback. The values are expressed in Hertz.

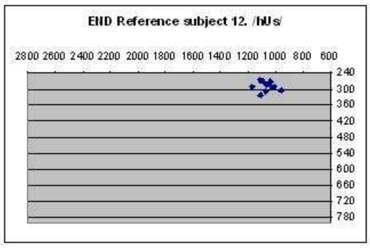


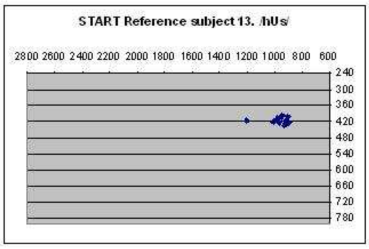


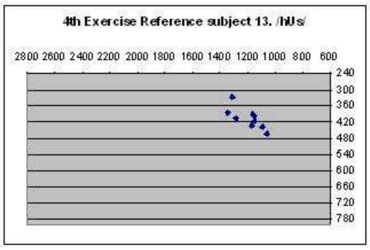


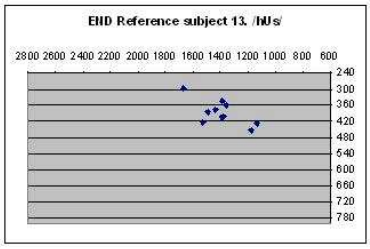


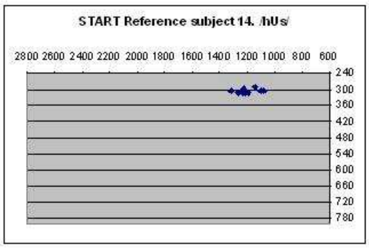


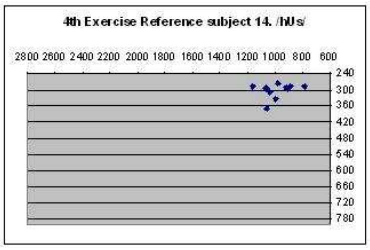


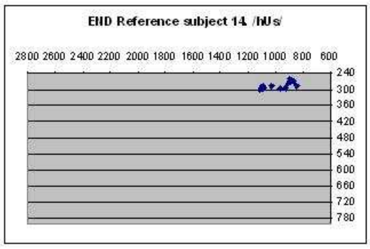


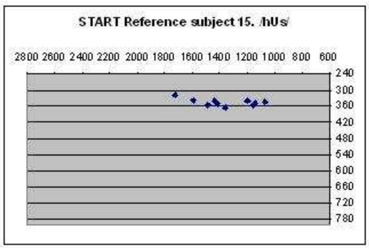


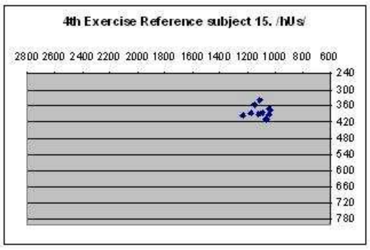


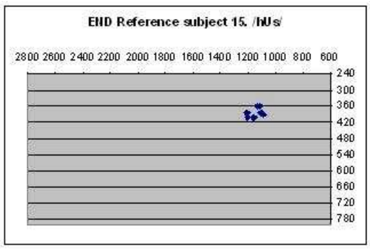


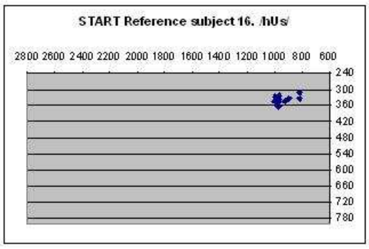


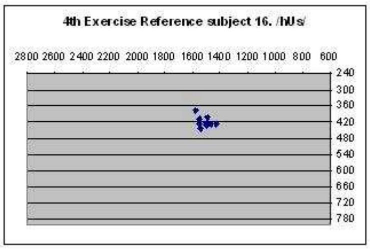


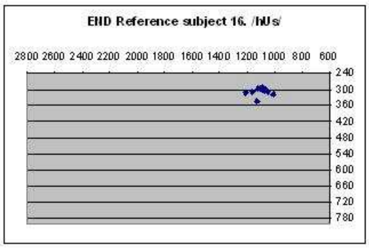


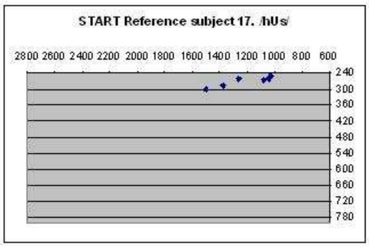


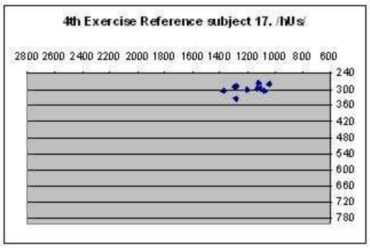


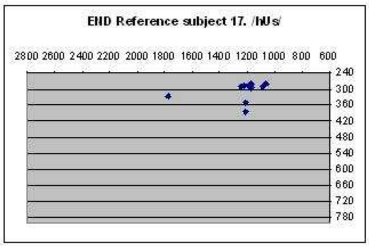

START Reference subject 16. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

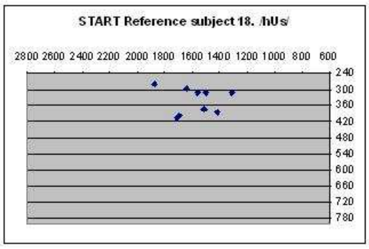

4th Exercise Reference subject 16. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

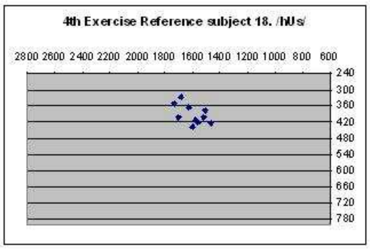

END Reference subject 16. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

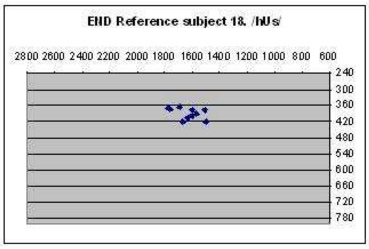

START Reference subject 17. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

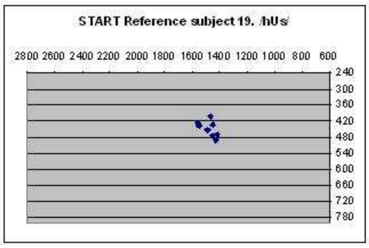

4th Exercise Reference subject 17. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

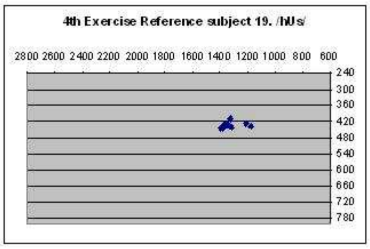

END Reference subject 17. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

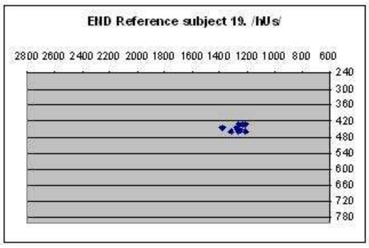

START Reference subject 18. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

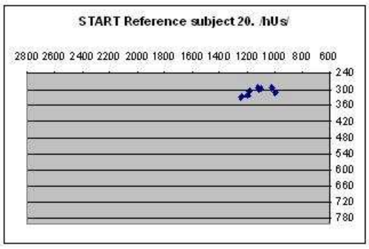

4th Exercise Reference subject 18. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

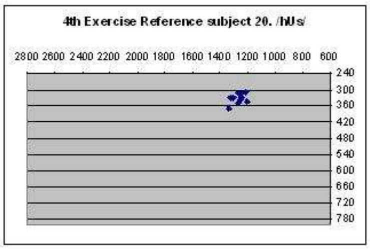

END Reference subject 18. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

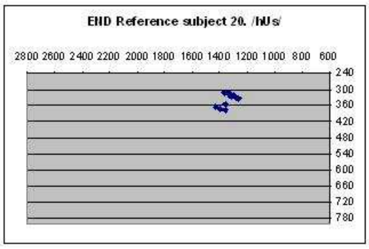

START Reference subject 19. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780


4th Exercise Reference subject 19. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780


END Reference subject 19. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780


START Reference subject 20. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780


4th Exercise Reference subject 20. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780


END Reference subject 20. /hUs/
2800 2600 2400 2200 2000 1800 1600 1400 1200 1000 800 600
240 300 360 420 480 540 600 660 720 780

Appendix I

Vokaalipelin testihenkilökaavake:

Nimi: ____

Ikä: __________

Ylioppilaaksitulovuosi: _________

Ylioppilaskirjoitusten paikkakunta: ____

	Kyllä:	Ei:	Lisätietoja:
Ruotsinkielinen tausta:	☐	☐	_____________________________________
Normaali kuulo:	☐	☐	_____________________________________
Normaali näkö:	☐	☐	_____________________________________
Normaali puhe:	☐	☐	_____________________________________

Testiosio:	PVM:	Kello:
ALKUTESTI + 1. harjoittelukerta:	_______________	_________________
2. harjoittelukerta:	_______________	_________________
3. harjoittelukerta:	_______________	_________________
4. harjoittelukerta + LOPPUTESTI:	_______________	_________________

____

Täytetään 4. laboratoriokäynnin jälkeen:

	PVM:
Elokuvalippu kuitattu:	_______________

____

Testihenkilön allekirjoitus

Appendix J

Vokaalipelin testauksen palautekaavake

<u>Kaavake täytetään kun kaikki 4 laboratoriokäyntiä on suoritettu.</u>

1) <u>Oliko testien ohjeistus mielestäsi selkeä?</u>

Erinomainen	Hyvä	Keskinkertainen	Heikko	Täysin surkea
☐	☐	☐	☐	☐

2) <u>Millaisena pidit testien kiinnostavuustasoa?</u>

Erinomaisena	Hyvänä	Keskinkertaisena	Heikkona	Täysin surkeana
☐	☐	☐	☐	☐

3) <u>Millaisena pidit palkkiota suhteessa käytettyyn aikaan?</u>

Erinomaisena	Hyvänä	Keskinkertaisena	Heikkona	Täysin surkeana
☐	☐	☐	☐	☐

4) <u>Testejä voisi kehittää/monipuolistaa/ohjeistaa paremmin.</u>

Täysin samaa mieltä	Jokseenkin samaa mieltä	En osaa sanoa	Hieman eri mieltä	Täysin eri mieltä
☐	☐	☐	☐	☐

5) <u>Opin testien kautta uutta.</u>

Täysin samaa mieltä	Jokseenkin samaa mieltä	En osaa sanoa	Hieman eri mieltä	Täysin eri mieltä
☐	☐	☐	☐	☐

KÄÄNTÖPUOLELLE VOI KIRJOITTAA KOMMENTTEJA/MUUTA PALAUTETTA TESTEISTÄ (esim. pelin käyttöliittymästä tms. testiin liittyvästä):

KOMMENTTEJA/PALAUTETTA: