An overview of a research project and preliminary results of two experiments on perception and production of foreign language sounds by musicians and non-musicians

Pastuszek-Lipinska, B.

journal: TMH-QPSR
volume: 46
number: 1
year: 2004
pages: 061-074

http://www.speech.kth.se/qpsr
An overview of a research project and preliminary results of two experiments on perception and production of foreign language sounds by musicians and non-musicians

Barbara Pastuszek-Lipinska
Department of Speech, Music and Hearing, KTH, Stockholm, Sweden, and School of English, Adam Mickiewicz University, Poznan, Poland
e-mail address: lipinska@speech.kth.se; tlumaczbpl@poczta.onet.pl

The paper has been accepted for a workshop "Speech perception within or outside of phonology?" - part of the 27th annual DGfS conference and will be presented on February 2005 in Cologne (Germany).

Abstract
This paper presents an overview of a new research project and outlines its design, goals and main aims, and the steps that have been already taken and will be undertaken within the project. Finally, preliminary results from two pilot experiments (pitch contour tracing and listening test) are presented. The project aims at investigating the influence of musical training and education on foreign language acquisition and specifically its influence on perception and production of foreign language sounds. This paper provides behavioural evidence on the issue. Studies and analysis of gathered data are currently underway and more detailed research results will be described in future publications.

1. Introduction
In a series of earlier experiments, a number of researchers have investigated foreign language speech perception and production by normal healthy people (for review see Best, 1995; Bohn, 1995; Strange, 1995; Hume & Johnson, 2003; and works by Escudero, 2002). However, there are still only a few investigations of the differences between people with normal hearing and those with a professionally trained auditory system – musicians.

Although music and language production (speech) are in many respects different they also have several similarities that may be inter-related. They are two specifically human phenomena. Both are conveyed by sounds and, in both, the human auditory system is involved. It is not surprising that such issues attract researchers from different domains to investigate the similarities and differences between music and language (for instance, music processing and language processing, and music perception and speech perception).

Our aim is to find out if one domain influences another and to what extent musical training, for example, can influence the process of second language (L2) acquisition.

1.1 What we know about language development?
Many research studies provide evidence that at birth, infants are able to distinguish all phonetic contrasts in the world’s languages (Polka, 1995). Several research study results have shown that young infants aged 6-8 months have excellent auditory capacities and can discriminate both native and non-native contrasts (Trehub, 1976). By the end of the first year of life, the ability seems to disappear (Kuhl, 1995; Werker & Tees, 1984), and gradually diminishes in the following years.

Researchers who have examined the evolution of language from the first months of a human’s life have reported that the ability to discriminate can be connected with auditory sensitivity (Kuhl, 2000) and that this capability is changed by speech produced by adults who have contact with an infant. In this way, language experience and exposure has been recognized as the most important agent that
provokes changes necessary to discriminate new speech sounds. Interestingly, the auditory acquisition and thus the prosodic acquisition of a particular language seems to be completed before a baby starts talking (first words by the end of the first year of its life) and all without any formal instructions.

There is a consensus that language experience plays an important role and has a crucial impact on speech perception and production (Gandour et al., 1998; Zhang et al., 2000). Many approaches to L2 perception suggest that first language (L1) background plays an important role also in L2 perception and has influence on the way in which non-native sounds are perceived (Ingram & Park, 1997).

There is a consensus that children are able to learn a particular language by only listening to its production, although during the process of language acquisition they have to detect the relevant acoustic cues and then the relative importance that the cues have in speech production (Nittrouer, 2000).

In speech perception people, i.e., native speakers usually are not conscious of the acoustic cues connected with their mother tongue.

Studies in second language learning have also found that listeners are more efficient at perceiving sounds of their mother tongue than those of a new language acquired later in life (e.g. Best et al. 1988; Best, 1994; Dupoux et al., 1997; Harnsberger, 2001; Hume & Johnson, 2003; Polka & Werker, 1994; Strange, 1995). It seems, however, that first language has an impact on second language and the process of a perceptual attunement to the native language contrasts decreases the ability of discrimination of some non-native contrasts.

People learning new languages in adolescence or adulthood often have difficulty in discrimination of non-native contrasts; the prosodic acquisition of a new language seems to be difficult (Werker & Tees, 2002).

Adults who start learning a new language, and already know in practice at least their native language, have to face many boundaries, because the process of maturation of their auditory system is completed (Rauschecker, 2003). Yet, research studies show that adults compare poorly with children in phonetic units discrimination.

Studies of adult subjects have shown, however, that the auditory ability to discriminate non-native contrasts and speech sounds still exists throughout the human life span but is subject to “reorganization” (Diamond et al., 1994; Polka, 1995; Werker & Pegg, 1992). So, it seems that the human capacity to discriminate foreign language sounds exists but is limited by lack of experience and stimulation during many years of life.

In support of this evidence, several laboratory studies have recently shown that perception of non-native speech sounds can be learned through special training (Dziubalska-Kolaczyn, 2003; Kjellin, 1999; Lively et al., 1994; Sari, 1996) and that the human auditory cortex can be shaped by the manner in which sounds are combined into words and sentences (phonological grammar) (Jacquemot et al., 2003) that support neuropsychological evidence on plasticity in the brain and human auditory functions.

Thus, adults have to work harder than children on acquiring a new language but, according to new research results, adults like children can achieve native-like production of foreign language sounds.

As scientists from other domains interested in cognitive science say, we can agree that “earlier means better” but we should also like to say that to enable appropriate stimulation is to enable an environment for changes, possibilities and new challenges.

Researchers also emphasize the important role of prosody in second language speech perception and production (Chun, 1998; Kjellin, 1999). They point out that when students start to learn a new language, some time is usually needed in order to learn to pronounce phonemes that are not present in their native language. Yet experience shows that a person with a good organization of a discourse, good grammar, and good knowledge of words and their meaning who lacks correct timing and pitch will be harder to understand.

So, it seems that there is a need to examine which agents can aid the process.

It would be also interesting to find out under which conditions human ability to discriminate different language sounds is maintained and trained across the life span.

---

1 Phonological grammar (Kay, JD 1989, Phonology: a cognitive view: LEA: Hillsdale, NJ.)
1.2 Music in human life and cognition as stimuli in treatment and therapy

Music plays an important role in all cultures and, similar to language, is a means of communication between people. Music accompanies our everyday life and several research studies have been so far devoted to the investigation of its influence on several aspects of human life.

Music is already used in different kinds of trainings that aim to improve health, e.g. after brain injuries, during mental and psychological trainings.

As a good example, results of research studies at the Institute for Music Research, University of Texas and the Research Imaging Center, University of Texas Health Science Center may be cited, where researchers examined the influence of melodic intonation therapy on fluency of speech after brain injury and reported very promising results.

Researchers from the Ohio State University have newly examined how patients complete a verbal fluency test while cardiac rehabilitation. Music is used as stimuli and researchers have reported significant improvement. Emery suggested, “listening to music may influence cognitive function through different pathways in the brain” (Emery et al., 2003).

Music is also used in foreign language classrooms as a means to improve, for example, verbal memory and intonation (Chan et al., 1998).

If short-time exposure (only listening conditions) can provide such a positive effect it would not be perhaps useless to ask about a long-time exposure (e.g. musical training and education).

So, in our study we are particularly interested in healthy adults and their perception and production of sounds of a new language, and in the influence of long-term stimulation, such as musical training, on second language acquisition.

1.3 What we know about the music impact on the language development?

Can we find any impact on language of music at the stage of the first language acquisition? Let us start from the beginning. As we have already said children seem to be masters of sound perception, but there is still the question of why and how the process proceeds and which components are crucial in the process.

Researchers working in speech perception and who are interested in speech development have found that infant-directed speech is in a number of aspects music-like (e.g. regular rhythms, slow tempo, pitch contours expanded and repeated with altered lexical or segmental content and varying tempo, extended vowels (as in song)) and finally musical qualities. Linguists also claim that infant-directed speech has a distinct suprasegmental structure (for review see: Trehub & Trainor, 1993).

Fernald (1992) even claimed that: “long before they understand language, babies find out what’s on their mothers’ minds through emotional clues in the parent’s exaggerated, musical speech”. So, according to this view, the musical components seem to be crucial in language acquisition and development.

Thus, on the basis of this consideration it is valid to ask whether these musical features of infant-directed speech are crucial for language development, and whether musical expertise is a means that helps in L2 sound perception and production. In other words, to what extent is musical stimulation a factor that influences and helps language acquisition?

1.4 Musical education and training – overall characteristic

Music education is a complex process that involves many different dimensions. It is based among other things on development of specific auditory functions. Students of music are trained in pitch discrimination, rhythm and timing accuracy. The most important part of this kind of education is to focus students’ attention on sounds. The human auditory system and memory are involved in this process. Emotions that musicians would like to show, and/or generate in the audience, ensure the final shape of the music performance.

1.5 Does music make changes? - Neurological evidence

A number of researchers provided evidence on the sensitive period both in music and in language. Researchers from the neurological domain have gained insight into the human brain, due to newly available technologies, and have provided evidence that human brain is plastic during all life-span (Syka, 2002).

Works investigating similarities and differences between musicians and nonmusicians at the level of the brain are currently very
advanced and show important differences (both qualitative and quantitative) between the two groups of subjects (for review see Peretz, 2003; and Schlaug, 2003). The brains of musicians have been recognized as possessing “exceptional neuroplasticity” (Munte et al., 2002).

However, only a few studies have been devoted to the investigation of differences and similarities between musicians and non-musicians at a behavioural level.

If at the first stage of L1 acquisition, musical cues of speech that enable language acquisition play an important role. So perhaps musical training that provokes qualitative changes in the human brain facilitates discrimination of foreign language.

It seems to be interesting to examine also behavioural changes that may follow the changes mentioned above (Peretz, 2003 etc).

In order to investigate the question of behavioural changes, we developed a research project and carried out two experiments.

2. Research design and methodology

2.1 Research question and hypothesis:

Question: Do musicians meet fewer problems with discrimination of non-native speech sounds?

Hypothesis: People with a musical background will have fewer difficulties with perception and production of new speech sounds as a result of constant auditory stimulation during musical education and training.

2.2 Experimental design and procedure

2.2.1 The Corpus

82 sentences in 6 languages (English: American, British, Belgian Dutch, French, Italian, Spanish: European and from South America, and Japanese) have been synthesized for the corpus. The ScanSoft® RealSpeak™ application was used for the purpose. Among sentences were questions, statements and orders. All sentences were recorded on CD (3 times repeated with pauses between them).

2.2.2 Research data

1) Participants

Conducted in Poland, a group of 106 subjects, Polish nationals, with and without musical education and training and with and without different languages competence were examined.

All subjects were recruited in Lodz and Kutno areas and participated in the study after stated consent and on a voluntary basis.

All subjects were native-speakers of Polish aged from 15 to 69 years (mean 32). All subjects reported themselves to have normal hearing although some of them filled out in questionnaires that had some hearing related illnesses (e.g. otitis, other temporal impairments) in the past, also subjects advanced in age could have age-related hearing changes.

At this stage of design of the project we intended to have two groups: the first composed of non-musicians and the second composed of professional musicians (having at least reached the level of secondary music school which is normal in Poland after 10-12 years of education).

After gathering background data we noticed that our first division (of musical competence) did not sufficiently describe the subjects. For instance some professional musicians after 10-12 years of education are currently not active in music, and some subjects that claimed to be non-musicians had some musical experience at childhood; there is also a small group of subjects – non-musicians who even without any formal training perform as unprofessional musicians-amateurs.

It seems that there are still some subjects who cannot be classified according to our current division.

Finally, however, our subjects were divided into 4 groups:

1/ musicians with at least 8 years of musical education and still active in music; in this group has been also classified professional organist who started his professional career after 6 years of formal training,

2/ people active in music as amateurs – both with and without a musical background,

3/ subjects with some musical experience in childhood but not active in music,

4/ subjects without any musical background.

2) Procedure

Subjects were asked to repeat as accurately as they could some synthetic foreign-language sentences played by a CD player (Grundig) placed in a quiet area. No other information was given to the subjects.
Participants

- nonmusicians
- people with musical background but not active in music
- active amateurs
- musicians

Figure 1. Subjects

All sentences produced by the subjects were recorded with Sharp MD-MT200 portable recorder and UNITRA-Tonsil Microphone MCU-53 with a linear characteristic. All subjects also filled out questionnaires with information on their own sex, age, education (including start date of musical education and training and contact with particular languages), music exposure, occupation, job, interests, as well as health (we asked subjects to give information on the eventual hearing problems and all illnesses which could produce negative effects on hearing).

Subjects without musical background also participated in a simple test of musical competence and abilities (Pastuszek-Lipinska, 2003) prepared with the following tasks: 5 sounds to repeat, 4 words to sing, test of harmonic perception - 4 examples: 1 sound, an accord of three sounds with the middle sound to repeat, 2 sounds with the low sound to repeat and finally an accord of three sounds with the highest sound to repeat, comparison of two melodies with only small changes (both in rhythm and in pitch), comparison of two times produced short melody in major key and then in minor key, and 4 rhythms to reproduce (clap with hands).

The test of musical skills was based on standard entrance tests to music schools in Poland, so we can assume that all musicians were able to do this without any difficulties. The assumption is based on real-life cases, e.g. it is not possible to start and then continue musical education without passing the described test.

Figure 2. Results of the test of musical skills.
3) **Analysis of the data**

All the information gathered has been organized and memorized into a database. The next steps are analyses by means of statistical analysis and the results are plotted in a series of graphs and tables with the support of Analyze-it +General 1.71 for Microsoft Excel program.

3. **Experiments**

3.1 **Experiment 1 - Pitch contour tracing**

3.1.1 **Pitch contour – background information**

The American National Standards Institute (1973) defines pitch as “…the attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low”.

Pitch is one of the structural patterns that are present both in speech and music.

In speech, a pitch contour is a continuous melody-like pattern and is directly related to the vibration rate of the vocal cords tones, and acts as localised semantic and grammatical information.

This melodic contour in speech is also the perceptual correlate of fundamental frequency (F0) over time, commonly called “intonation”.

When we focus on speech production in the context of the current research project we cannot ignore the following findings and research results:

- First, on human beings ability to the “voluntary control over the precise acoustic patterns of vocal expression”. As well as on the fact that this ability plays a crucial role in both language and music, and particularly in speech production (Juslin & Laukka, 2003).
- Second, on the effect of superior auditory skills (such as in musicians) on vocal accuracy which reported that musicians even without vocal training showed greater vocal production accuracy (Amir et al., 2003).
- And third, that shown “the auditory-vocal system controlling voice F0 seems to operate mainly at the supra-segmental level, not at the syllabic level” (Natke et al., 2003).

Prosodic (suprasegmental) features of a given language have been very often said as being crucial for L2 acquisition. Thus, on the basis of the evidence we can claim that prosodic features (suprasegmental features) play crucial role in language acquisition and e.g. intonation – melody of a language express not only grammatical meaning (statement, question, order) but also emotions.

Yet, in a study by Patel et al. (1998) authors suggested that the same neural resources may be shared in the processing of F0 contours in speech and in melody contours in music and that perhaps there may be also similarities in pitch contours.

When we come back to language development and analyze the role of pitch contour processing we should take into account that pitch contour processing may be an important perceptual organizational device for infants, which helps them in organization of complex auditory patterns both in musical and speech sequences and particularly prosodic aspects of speech (pitch, stress, and junction patterns) as Trehub & Trainor (1993) suggested.

So, on the basis of the presented data it would be interesting to find out whether in foreign language people with long training and stimulation of auditory system through musical training and education perceive and produce foreign language sounds more appropriately (meaning closer to native-like production or to a given sample or at least more natural due to the stimulation and sounds sensitivity).

3.1.2 **Experiment design and procedure**

1) **Subjects**

Recordings with 106 speakers were analyzed. One person was excluded because he did not repeat the second word of the sentence. Due to the significant different sizes of the groups only two, the biggest and simultaneously the most important for our investigation, have been chosen for current results presentation: professional musicians (53) and nonmusicians (38).

2) **Material**

The research data was the question “May I help you?” (US English) produced by each speaker.

3) **Procedure**

In order to investigate similarities and differences between musicians and nonmusicians in pitch contour accuracy fundamental frequency (F0) of each word from the sentences produced by speakers have been extracted and analyzed.
Soundswell 4.0 program has been used in analysis of F0 of the question “May I help you?” produced by each speaker.

4) Data analysis

a. Results
Several analyses were performed:
1. Mean and max fundamental frequency (F0) of each word in the researched sentence.
2. Calculations of changes in mean F0 between syllables with semitones (in order to eliminate differences in frequencies between female and male voices), standard deviation from the sample recording and correlation with the sample.
3. And then data have been analyzed with statistical functions (SPSS and Analyzeit+General 1.71 for Microsoft Excel programs).

No significant differences between musicians and non-musicians’ production at the level of pitch contour were found.

Figure 3 shows results (mean, standard deviation and standard error tables) within two of the biggest groups of subjects: professional musicians and non-musicians. The graphs show standard deviations of the subjects’ production (mean F0 between syllables with semitones) from the original sample.

b. Discussion
The results achieved cannot be accepted as final because of some additional aspects. The first and very important aspect seems to be the described above role of experience in language acquisition.

English is a very popular language, which could be heard accidentally by all subjects so it seems that analysis of other languages or data with F0 manipulated could provide other results.

The second aspect relates to results of similar research project newly published by Schön et al. (2004). The authors examined whether extensive musical training facilitates pitch contour processing not only in music but also in language. Their results provide evidence that extensive musical training influences the perception of pitch contour in spoken language.

c. Conclusions
At this stage no significant differences in pitch contour accuracy between groups, and especially between two of the biggest groups (musicians and non-musicians), have been found. Both musicians and non-musicians did not have difficulties with pitch accuracy and intonation in the described sentence. Current research project will be repeated with: smaller group, natural stimuli and with languages unknown for all subjects.
3.2 Experiment 2 - Listening test

3.2.1 Experimental design and procedure

1. Subjects

Group of subjects consisted of 7 native speakers of English, namely 4 colleagues from KTH, 1 from Indiana University (US) and 2 from York University (England). 4 females and 3 males, aged from 31 to 62 (mean 49). All subjects participated as listeners in the current test on a voluntary basis.

2. Material

Research material was the sentence “May I help you?” repeated by 106 speakers – native speakers of Polish.

3. Procedure

In order to relate acoustical cues to human perception, listening tests are undertaken and different rating methods are used. Taking into account a large number of speakers in the current study, the Visual Analogue Scale (VAS) has been chosen as a verified means to measure a variety of subjective phenomena. A listening test was prepared and a computerized listening test (program Judge by Svante Granqvist) was run with a panel of listeners with the aim to investigate how native speakers of English perceive musicians’ and non-musicians’ production. The subjects’ tasks were to rate, for each stimulus, on a Visual Analogue Scale whether they felt that they heard bad or good English and put results on the scale with a cursor. The subjects were presented with a visual analogue rating scale on the computer display, the scale ranges from 0 marked ‘bad’ to 1000 marked ‘good’. Each subject was presented with a differently randomized list of stimuli. The program recorded all response settings. The test was preceded by a short introduction explaining that recorded subjects were native speakers of Polish who heard synthetic stimuli and after three repetitions reproduced heard sentences with taking into account both segmental (vowels and consonants) and suprasegmental (intonation, rhythm, stress, and rate) features. During the test, subjects could listen as many times as they needed to a given stimulus before giving a score. The task lasted around 20 minutes.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>95% CI of Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>NONMUSICIANS</td>
<td>39</td>
<td>341.370</td>
<td>163.369</td>
<td>26.160</td>
<td>288.412 to 394.328</td>
</tr>
<tr>
<td>PROFESSIONAL MUSICIANS</td>
<td>53</td>
<td>488.720</td>
<td>139.349</td>
<td>19.141</td>
<td>450.310 to 527.129</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Median</th>
<th>IQR</th>
<th>95% CI of Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>307.714</td>
<td>220.143</td>
<td>262.857 to 384.714</td>
</tr>
<tr>
<td>508.571</td>
<td>180.714</td>
<td>463.857 to 541.857</td>
</tr>
</tbody>
</table>

Figure 4. Graphs with t-test analysis for dependent samples illustrating the relation between grouping and the dependent sample (mean perceptual rating scores given by listeners).
4. Data analysis
In the current paper, however, all data have been involved due to the different sizes of the groups. When differences between the groups are presented, only two the biggest groups are involved (39 non-musicians and 53 professional musicians).

Data have been analyzed with SPSS and Analyze-it + General 1.71 for Microsoft Excel programs.

a. Results
1) We have found significant differences in musicians’ and non-musicians’ production of the researched question, “May I help you?” Listeners have perceived musicians as more fluent than non-musicians.

Figure 4 presents results achieved for these two of the biggest groups. Tables with descriptive comparative analyses and graphs between the groups are presented above.

According to listeners – native speakers of English’ scores the differences between the two groups are very significant.

2) We have also found significant correlation between musical abilities and average scores given by listeners.

Results of 2-tailed Pearson correlation test show that p-value (statistical significance) is 0.0016 and is highly significant. This may suggest that musical skills also play a role in the processes of speech perception and production.

Test – Pearson correlation

<table>
<thead>
<tr>
<th>n</th>
<th>106</th>
</tr>
</thead>
<tbody>
<tr>
<td>r statistic</td>
<td>0.30</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.12 to 0.47</td>
</tr>
<tr>
<td>2-tailed p</td>
<td>0.0016 (t approx.)</td>
</tr>
</tbody>
</table>

Pearson correlation was performed in order to measure relationship between mean rating by listeners (average), and musical skills (taking into account perceptual rating scale (range from 0 to 1)).

Figure 5. Pearson correlation results between mean production according to listeners’ scores and musical skills.
3) We have also found correlation between subjects’ production and language experience.

Results of 2-tailed Pearson correlation test presented in Figure 6 that measured correlation between mean rating by listeners (average) and language experience also show significant p-value that is 0.0001.

Test - Pearson correlation

<table>
<thead>
<tr>
<th>n</th>
<th>106</th>
</tr>
</thead>
<tbody>
<tr>
<td>r statistic</td>
<td>0.36</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.18 to 0.52</td>
</tr>
<tr>
<td>2-tailed p</td>
<td>0.0001 (t approx.)</td>
</tr>
</tbody>
</table>

**Figure 6.** Pearson correlation - results showing correlation between English experience and mean rating.

**Figure 7.** Graphs with comparative analysis of listeners’ mean rating scores including all subjects (speakers).
Figure 7 presents scores given by all participants of listening test. The scores differ between subjects and it seems that their personal level of acceptability plays an important role in the evaluation process.

In order to evaluate reliability of obtained data we used Cronbach’s Alpha (a tool that is used for assessing the reliability of scales, in other words for assessing the consistency of received data).

When data have multidimensional structure this parameter is usually low. In our study however Cronbach’s Alpha on standardized items is very high

$$\alpha=0.839$$

On the basis of this parameter we can see that the results are very consistent and all listeners’ scores have unidimensional structure.

4. General discussion – on experimental data and procedure

In recent years, several tests have been carried out in different conditions: e.g. only hearing, visual and audiovisual conditions.

Experiments have also been carried out that involved a speaker’s face and accompanying gestures as well as speech sounds (Traunmüller, 1998; Massaro, 2001).

The current study used only the auditory signal as a perceptual cue. The aim of the study is to examine behavioural response of musicians and non-musicians on foreign language sounds.

4.1 Subjects and their attitude

Although all subjects gave their consent to the participation in the project their dispositions were different. Several subjects were nervous of this kind of experiment, some of them were convinced that their performance would be negative, others were doubtful and almost convinced that they cannot be useful as a subject. Some of our subjects claimed that the task was too difficult for them.

We think it is an important element that could influence the results negatively because motivation plays an important role in performance.

4.2 Synthetic stimuli versus natural stimuli

In the current study, synthesized sentences generated by a text-to-speech system were used as stimuli. Although that kind of stimuli is now very popular, and a number of researchers have already utilized computer-synthesized stimuli as research material, there are, however, still some doubts about the influence of unnatural speech sounds on the results. Also in question is the influence of that kind of stimuli intelligibility.

Researchers have conducted several research projects in order to analyze perception of synthetic versus natural stimuli and, on the basis of the data; it is known that there are several advantages (e.g. the possibility to analyze separately a given parameter, to manipulate F0) but also several disadvantages (e.g. the above-mentioned artificiality of the sounds).

Bailly (2003) carried out several experiments that had as a main goal to compare natural versus synthetic speech in shadowing speech (meaning imitated just after heard stimuli).

He observed that an average delay (time needed to start imitate speech) is different for natural stimuli (less than 50 ms) than for synthetic speech (exceeds 100 ms); he claimed that this could be mainly due to prosody differences and that actual text-to-speech systems still generate “inappropriate and impoverished prosody”.

Before the samples were generated for the purpose of the current research project a number of text-to-speech systems were examined and the ScanSoft® RealSpeak™ application was chosen as the most intelligible (in our opinion) system, even though some sentences still sound artificial.

But yet, only a few of our subjects realized that they heard non-natural voices.

None of our subjects was familiar with speech synthesis.

So, it seems to be important to note that in the listening experiment reported below, our subjects in fact repeated the imitations of synthetic stimuli and not natural human voices. This may have influenced our results.

We can claim that subjects – both musicians and non-musicians have the same starting point but the question as to whether and how that might influence results still exists. So, the

---

3 See Corpus
experiment will be repeated in the future also with natural stimuli.

4.3 Work with a microphone

Another problem is connected with set-up of a microphone. Not all people are familiar with microphones and sometimes, even after clear instructions that the distance between the microphone and the mouth should remain unaltered throughout the experiment (in this case 20 centimetres), the subjects attempt to move the microphone. As a result, we obtained some recordings that were either too loud or too soft. In the future, a fixed distance will be ensured.

Such difficulties in recording are well known to researchers working in the domain of sound analysis; we intend here, however, to show all potential risks and weaknesses of our research results, as well as to highlight how we shall introduce all necessary changes in the future research studies.

4.4 Using VAS and data reliability

The next doubt is connected with applied scale itself, which provoked difficulties and could be a means of potential mistakes and inconsistencies. The number of samples is very big and it seems to be a difficult task to precisely evaluate them.

Listeners claimed after the test that after they had evaluated a given subject as very bad (e.g. 0) then in some cases they found other subjects even worst but they did not have a scale for lower evaluation. One of the listeners (subject F) used only a half of the VAS scale.

In order to avoid in the future problems with using the VAS by listeners (see page 11), information on the whole scale use will be added to the instructions before the next listening tests.

In order to eliminate the problems described above in the future the following suggestions and remarks of listeners will be introduced:

- List of instructions not only oral but also on paper so that to ensure they are the same for each listener.
- More detailed information about rating range – it seems that scale from bad to good is not enough, so in the next listening test the two definitions will be additionally described: “barely understandable” and “almost native-like production”, so that to help listeners in evaluation process.
- Training of listeners before starting whole test – showing how big differences between subjects are possible (6-10 the most varied samples will be presented at the beginning).

5. Conclusions

Steps undertaken so far and described above could be affected by some weaknesses, so it seems reasonable to repeat some of the experiments and to introduce some changes in the procedure. It would be also advantageous to take into consideration all suggestions by colleagues who participated in the project and who gave their opinions about the proposed approach.

Research results have shown some directions on aspects that should be re-examined and reanalyzed in order to provide data as reliable as possible.

The preliminary results of the experiments show evidence that in some way confirms the hypothesis on positive influence of musical training and education on second language acquisition (SLA) and particularly on foreign language sounds perception and production. However, there is still a need to introduce some changes and improvements to the methodology in order to receive more reliable research results as well as to continue in depth analysis of all gathered data. Although the evidence is small it is, at this point, worth pursuing the question of whether there is an influence of musical training on second language acquisition. We will thus repeat experiments with other languages and develop some new experiments with natural stimuli. We also intend to observe our two groups while spontaneous speech (when subjects know a given language) and after some training period, so that to achieve complex data on how perception and production of foreign language sounds is influenced by musical education and training.

6. Acknowledgements

These investigations are supported by EU Marie Curie Fellowship contract No. HPMT-CT-2000-00119. The author would like to thank all participants for their kind participation all remarks and patience. To Dr. Peta Sjölander, Dr. Svante Granqvist, Prof. Johan Sundberg and Prof. Robert McAllister for support and valuable
discussions, and to Dr. Joakim Westerlund for help with the statistical analyses.

7. References


Fernald A (1992). Speech on the annual meeting of the American Association for the Advancement of Science (AAAS) in Chicago on Feb. 8.


