euronoise

Influence of music education on second language acquisition

Barbara Pastuszek-Lipinska

Adam Mickiewicz University, al.Niepodleglosci 4, 61-874 Poznan, Poland
energin@wp.pl

To explore the extent to which music education influences second-language acquisition, two groups of native Polish speakers, musicians and non-musicians, were asked to reproduce sentences in six languages: English, French, Italian, Spanish, Japanese, and Belgian Dutch. The speech stimuli were developed with a text-to-speech application and differed phonemically, phonostylistically, and in length. The paper includes results of a general auditory analysis of subjects' productions as well as the results of a web-based listening test with a panel of native speakers of the involved languages. All collected data were also analyzed with statistical tools. The results revealed that music education exerted a measurable impact on speech perception and production. Musicians outperformed non-musicians in the study. From the results, it appears that the influence of musical expertise extends beyond music processing to speech processing, and the strength of this influence is connected not only to auditory training. Therefore, the superior performance of the musicians in the task may be interpreted as evidence that music education is an enabling factor in the successful acquisition of a second language. It also indicates that the impact of musical training is not a myth, but has a scientific basis.

1 Introduction

The most fundamental questions in language and language acquisition research relates firstly to the extent to which language development relies on innate mechanisms or predispositions, and secondly to the role of exposure and experience in successful language acquisition. The present paper is aimed mainly at examining the hypothesis that music education and training may exert a positive impact on the process of second language acquisition. Thus, it is assumed that experience and interface between different disciplines play a role.

The goals of the study were to examine the behavioural response of musicians and non-musicians to stimuli composed of foreign language sounds and their structures as well as to observe the effects of special auditory training on the realization of the task. Taking into consideration the main assumptions in the current study, only auditory signals have been applied as a perceptual cue, so that to observe how musically trained and untrained subjects cope with different types of auditorily provided speech material.

2 Research design and procedure

2.1 Research corpus

82 word sequences in 6 languages (English: American (15), British English (14), Belgian Dutch (10), French (10), Italian (10), Spanish: European (6) and South American (4) and Japanese (10)) were synthesized for the corpus.
The ScanSoft® RealSpeak™ application was used for this purpose. Languages were chosen according to their typological classification; among them there were stimuli that included stress-timed, syllable-timed, and morae-timed languages. Among the sequences were questions, statements, and orders. The corpus also contains some phonological words, names, and/or other short word sequences. Thus, the stimuli differed phonemically and phonostylistically and contained a variety of lexical items; the length of the sequences was diversified as well.
All word sequences were recorded on CD, and were repeated three times each, with short gaps left between the repetitions of each sequence and a longer pause after each sequence that provided speakers with time needed to repeat the sentence. In this way, a recorded corpus was developed, which served for further data collection

2.2 Participants

A group of 106 subjects was examined: all of the participants were native speakers of Polish, but the participants had varying levels of language competence, and some had had musical education and training while others had not. All subjects participated in the study on a voluntary basis. They were not paid for their participation in the study. All subjects were aged from 15 to 69 years, with a mean age of 32 (median 28). All subjects reported that they had normal hearing.

2.3 Questionnaire

For the purpose of the study, a special questionnaire was developed. The questionnaire was designed to elicit information on each participant's sex, age, education (including the start date of their musical education and training, as well as their contact with foreign languages), music exposure, occupation, job, interests, and health (subjects were asked to give information on previous hearing problems and all illnesses that could have a negative impact on their hearing).

Although prospective participants were informed prior to the study that the main criterion of participation in the procedure was musicianship, several inconsistencies and instances of contradictory data were noticed during data analysis.

In both reported experiments, speakers were divided into two groups. The main criterion for this classification was the period of music education. Musicians had to be actively involved in musical training during 8 and more years at the moment of examination. This means that the group of musicians contained 53 subjects. All subjects who had had less that 8 years of music background and who had not been active in the field were included in the group of non-musicians (53).

2.4 Main procedure

In order to examine and compare how subjects with musical training and those without musical background perceive foreign-languages constructs and sounds, their imitation of those constructs and sounds was tested. The task was meant to examine the participants' ability to integrate different components of linguistic information, such as phonology, syntax, and intonation [1]; the presented study was also similar to the study conducted by Baddeley and his collaborators, which utilized nonsense words and immediate serial recall [2].

The task was not a pure measure of the enumerated components, but was instead aimed at finding a key to success or failure in the acquisition of language sounds and structures (both in terms of perception and production). Subjects were asked to repeat, as accurately as they could, some synthetic foreign-language word sequences played on

a CD player (Grundig) placed in a quiet area. No other information was given to the subjects. Examinees were not informed that they had heard synthetic stimuli. All sentences produced by the subjects were recorded with a Sharp MD-MT200 portable recorder and a UNITRA-Tonsil Microphone MCU-53 with a linear characteristic. The data were collected in different areas, not in a laboratory, which was not available to the author; thus, the prepared technical equipment enabled the author to move about easily and reach the subjects in different places, even at their homes.

Using these tools, more that 10 hours, and a half of recorded material were obtained. The results of the test of musical ability, as well as detailed information on subjects, were gathered on paper. Speakers were asked to repeat foreign language word sequences, so that the researcher might examine which language components and levels were most problematic for subjects and whether the differences between the two groups of subjects (musicians and non-musicians) could be noticed in such a task. The data were gathered with the assumption that all normal healthy humans possess the ability to imitate sounds easily [3, 4].

2.5 Test of musical skills

In order to gather data on the musical skills of the participants, a special test designed to examine musical abilities was developed. The test was not a standardized test, but it was developed so as to examine general musical skills and memory for music stimuli in a short time. Standardized tests were not chosen due to their length, which was not acceptable in this case.

The observation gained from the pre-test procedure [5] showed that the whole examination could not last longer than 30 minutes, because it was a very boring and exhausting task. Thus, subjects without any musical background participated in a test of musical competence and abilities [6].

The test was prepared with the following tasks: participants were asked to repeat five sounds and sing four words according to the model presented on a CD. They were also asked to respond to four sets of sounds and chords: a sound, a chord of three musical sounds with the middle sound to sing, two sounds with the lower sound to sing and finally a chord of three sounds with the highest sound to sing. Participants were then asked to compare two melodies that were slightly different in rhythm and in pitch, to compare two versions of a short melody produced in a major key and then in a minor key, and then to reproduce four rhythms by clapping their hands (see Figure 1).

All the tasks were recorded with Sharp MD-MT200 portable recorder and UNITRA-Tonsil Microphone MCU-53 with a linear characteristic, and then the author developed a CD with the tasks and recorded instructions.

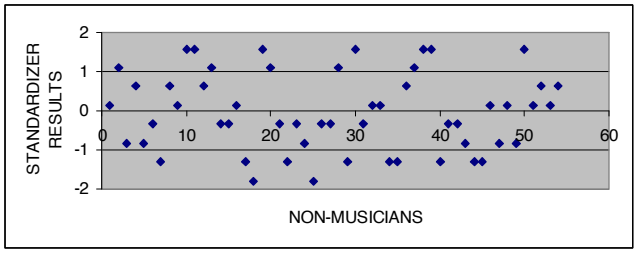

Fig. 1 Results of the test of musical skills and memory for musical stimuli in non-musicians.

The task lasted around 5 minutes. The test of musical skills was based on the standard entrance tests to music schools in Poland; it contained similar tasks to those that are included in that standardized tests of musical skills, but the number of questions was limited. On the basis of the pre-test results, it was assumed that all musicians were able to pass the tasks without any problems. The assumption was also based on real-life cases – namely, it is not possible to start and then continue one's musical education without successful completion of the described test.

Non-musicians' responses to musical stimuli were not recorded. The present author (she is a professional musician) rated their productions auditorily; she used a three-grade scale to evaluate four abilities – pitch tracking, rhythmic skills, harmonic hearing, and memory for music stimuli. Results were noted in questionnaires that had been earlier prepared separately for each participant. It can be seen that non-musicians' performances differed significantly among them.

3 Experiment 1

In order to examine how musicians and non-musicians responded to synthetic stimuli, all recordings were listened to and analysed. The main goal was to determine whether all subjects performed at an identical or similar level. It was assumed that there might be differences among subjects (and statistically among groups). It was also assumed that subjects' performances might differ between languages, due to their typological differences.

The study did not aim to ascertain solely whether musicians repeated word sequences better than non-musicians, also aimed to determine which aspects or components of language caused both groups the greatest difficulties. Another aim was to observe whether accuracy at the global level accompanied with accuracy at the local level. In order to discover the exact differences in the mispronunciations, all the word sequences were analyzed. Moreover, detailed analyses are provided, along with transcriptions of the original sentences in English (both American and British) their production, and a list of the most frequently occurring errors and mispronunciations.

3.1 Research material

Digitized recordings with 106 speakers producing 82 word sequences in six languages: English (American and British), French, Belgian Dutch, Spanish (European and South American), Italian, and Japanese. Recordings of 15 sentences in American English and 15 sentences in British English reproduced by 106 participants were subjected to more detailed analyses.

3.2 Procedure

Given the previous observations on the most common difficulties that Polish native speakers encounter in learning foreign languages (i.e. quality and length of vowels, perception and production of consonants and consonantal clusters, articulatory and pronunciation problems, and memory problems), it was recognized that there were many aspects that might cause inconsistencies.

Digitized English speech material was subjected to a more detailed analysis: namely, the material was treated at two

levels (the local-segmental and global-suprasegmental) [7, 8, 9]. Most attention was given to word and inter-word phonetics, but attention was also paid to other aspects of phonetics, such as the quality and length of vowels, the quality of consonants and consonantal clusters, and the details of mispronounced words and prosody, including rhythm and accents distribution as well as overall intonation, and these elements were discussed within the phonetic analysis of English speech material [7, 10].

3.3 General data analysis

The author rated the speech samples by an impressionistic auditory analysis. In the first round of data analysis, the scoring procedure was based on a general review and observation of whether all speakers responded to the stimuli and were able to repeat the speech material in the given time and with appropriate accuracy. Almost all subjects encountered difficulty with at least one sentence.

In order to evaluate the difficulty of the task, the Difficulty Factor was calculated, in which the optimal level equals 0.5 and which is usually used to check the proportion of respondents who were able to give the right answer to a given question or task.
The difficulty factor may be calculated using the following formula:

D = c / n
D - Difficulty factor
c - Number of correct answers
n - Number of respondents

As the purpose of the current study was to discriminate between different levels of performance, items with difficulty values between 0.3 and 0.7 would be most effective in making that discrimination. The difficulty factor in the current study shows that the applied procedure and its difficulty were close to optimum, and the task was feasible. Namely, the factor equals 0.56 in case of musicians and 0.39 for non-musicians, which means that the task was available for both groups of speakers.

3.4 Conclusion

Musicians produced more sentences, and it can be assumed that they encountered fewer difficulties with the task. There was only one sentence for which the number of correct productions in the musicians' group was the same as the number in the non-musicians' group. Furthermore, the gathered data may suggest that the musicians had better memories, and that this parameter enabled them to perform better. They encountered fewer difficulties in remembering speech passages, and thus it may be assumed that they encountered fewer difficulties with the task.

4 Experiment 2

To observe how participants were perceived by native speakers of all languages involved in the study, a web-based experiment was designed and conducted. The number of participants-raters was unfortunately not representative for all the languages used in the study.

4.1 Participants

The group of raters consisted of twenty four native speakers of American English: sixteen females and eight males, aged 18 to 68 (mean 33.2); two native speakers of Belgian Dutch: one female and one male, aged 23 and 41 (mean 32); four native speakers of British English: one female and three males, aged 23 to 60 (mean 41), fifteen native speakers of French: seven females and eight males, aged 21 to 41 (mean 29.26); eight native speakers of Italian: four females and four males, aged 22 to 39 (mean 29.25); three native speakers of Japanese: two females and one male, aged 24 to 57 (mean 37.66); and nine native speakers of Spanish: eight females and one male, aged 23 to 51 (mean 36.2).

Participants-raters were recruited through Info Childes, Auditory, Speech Prosody, and Topica lists and services, as well as through personal contacts with scientists working at foreign universities. All subjects participated as listeners in the current test on a voluntary basis. They were informed about the procedure as they could read a file with detailed instructions before running the test procedure. The instructions were prepared with the aim of guiding participants through the whole experimental procedure without personal assistance from the researcher.

4.2 Material

Research material comprised word sequences repeated by 106 speakers who were native Polish speakers. There were seven word sequences, one from each language (see Table 1 below for chosen word sequences).

4.3 Procedure

To carry out an internet-based study, special software was developed. The Visual Analogue Scale (VAS) was used to rate the productions of the speakers.

A web-based listening test was prepared and a computerized listening test (*the Ocen-wav program*) was run with a panel of listeners to investigate how native speakers of all included languages perceive musicians' and non-musicians' production.
The listeners' tasks were, for each stimulus, to rate on a Visual Analogue Scale whether they felt that they heard bad or good production of their native language and to put their results on the scale with the cursor. Before starting the real test, they were encouraged to listen to some samples of extreme productions to have a view of how different the productions might be. Moreover, they could listen to the synthetic samples chosen from the corpus and heard by the speakers whose productions the listeners were asked to rate. The test was preceded by a short written introduction in which it was explained that the recorded subjects were native speakers of Polish who heard synthetic stimuli and who, after three repetitions, reproduced the heard sentences, taking into account both segmental (vowels and consonants) and suprasegmental (intonation, rhythm, stress, and rate) features. During the test, the subjects could listen to a given stimulus as many times as they needed before giving a score. No information on the recorded subjects or their musical background was provided. The whole procedure lasted around 20 minutes.

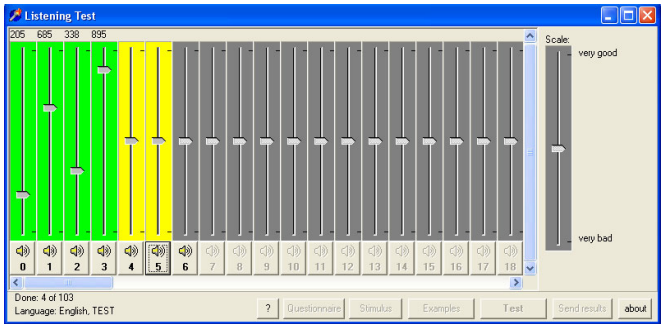

Fig. 2 The display of the ocen-wav program used for the web-based listening test.

The subjects-listeners were presented with a Visual Analogue rating Scale on the computer display (Fig. 2); the scale ranged from 0 (*barely understandable*) to 1000 (*almost native-like production*). Each subject was presented with a differently randomized list of stimuli. The program did not allow the listeners to close the procedure before rating all the provided productions, and it recorded all response settings. At the end of the procedure, the listeners pressed the *submit* button to send the results of the test and questionnaire to the email address of the author of the study. Moreover, the scores were sent anonymously.

Several speech samples from each language were excluded due to the lack of whole sentence, a word or other part of the production of the repeated sample. In several cases, files were excluded due to the weak quality of a sample. As a result, the raters scored: 103 productions of American English (52 musicians and 51 non-musicians), 102 Belgian Dutch (49 musicians and 53 non-musicians), 106 British English (53 musicians and 53 non-musicians), 104 of French (52 musicians and 52 non-musicians), 98 Italian (48 musicians and 50 non-musicians), 104 Japanese (52 musicians and 52 non-musicians) and 98 Spanish (49 musicians and 49 non-musicians).

4.4 Data analysis and results

Data have been analyzed with Analyze-it + General 1.73 for Microsoft Excel and Microsoft Excel programs. Each aspect of the language data has been treated separately before the results were consolidated and compared.

SEQUENCE	MUSICIANS		NON-	
	MEAN	SD-M	MEAN	SD-M
Sorry to keep you waiting.	547.90	0.15	463.34	-0.16
Een fantastisch spektakel.	462.09	0.22	357.96	-0.20
Is it yours?	530.03	0.11	489.04	-0.11
Tout le monde!	520.98	0.11	466.60	-0.11
La storia si ripete.	562.51	0.16	475.36	-0.15
Konnichiwa.	491.31	0.08	459.77	-0.08
Más vale tarde que nunca.	429.91	0.20	310.56	-0.20

Table 1 Results of the web-based listening test

There were significant differences between the musicians' and non-musicians' productions of the involved word sequences in all languages. The difference between the numbers of speakers was not important, as the applied classification allowed the involvement of all participants. Raters scored the musicians' productions as more fluent than the non-musicians'. Table 1 (below) shows the mean results of the two groups scored by native speakers.

According to raters' scores, there were differences between the two groups of speakers. Figure 3 (below) presents mean results of the listening test.

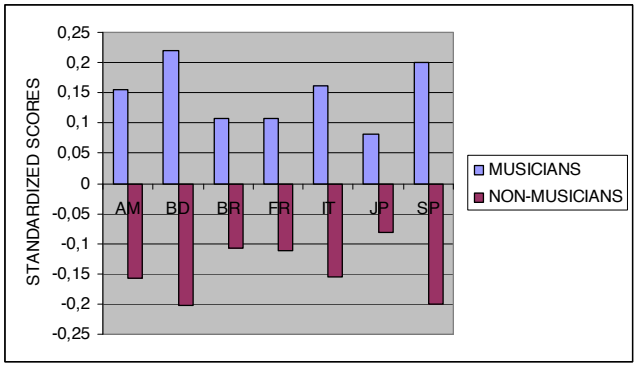

Fig. 3 Mean standardized scores obtained by two groups of examinees in cross-linguistic listening test.

Legend: AM-American English, BD-Belgian Dutch, BR-British English, FR-French, IT-Italian, JP-Japanese, ESP-European Spanish, ASP- South American Spanish.

The graphs above present the standardized scores given by the native speakers; it can be seen that in all examined sentences, musicians obtained higher scores (above the average at 0 level), while non-musicians obtained lower scores (below the average).

LANGUAGE	CRONBACH'S ALPHA
American English	0.977
Belgian Dutch	0.711
British English	0.847
French	0.954
Italian	0.945
Japanese	0.833
Spanish	0.968

Table 2 Cronbach's Alpha for all word sequences used in the study.

Cronbach's Alpha was applied to verify the reliability of the obtained data. In this experiment, in almost all languages the coefficient is larger than .70. To sum up, it may be stated that all native-speakers of the involved languages perceived musicians to be more fluent than non-musicians.

5 Recapitulation of results of the two experiments

The current study, which was aimed at investigating the influence of music education on second language acquisition, included several steps. The first step investigated participants' musical skills and memory for music sequences. The second step examined how musicians and non-musicians tackled foreign language word sequences.

It should be noted that the successful realisation of the task, which consisted of shadowing repetitions, has been recognised as a good indicator of phonological short-term

memory. This, in turn, has been recognised as a predictor of language learning success [11]. The digitized productions of the participants were analysed and examined, using several different tests and experiments so as to obtain a view of how musicians' and non-musicians' productions differ; some data that may be relevant in answering this question are posited in the current paper.

To ascertain whether the scores given by the author in the general analysis of the whole database correlate with the scores given by native speakers, the results were compared and a Pearson's correlation was conducted. The results proved that the scores correlate and have a unidimensional structure.

There was a significant correlation between the results of the general analysis and the web-based listening test in all languages used in the study.

The results of the test revealed that the scores given by the author and those given by native speakers were strongly correlated and had a very high probability level ($p<.000$).

The test showed that the differences in results among groups may be interpreted as statistically significant, because the musicians' mean standardized score was above the average of 0, equalling 0.145. In turn, the mean standardized score for non-musicians is below the average, equalling -0.141. The standard deviation in the case of musicians is lower (0.4575), and in non-musicians, higher (0.5193), which suggests that the second group was less homogeneous and their results were more distributed.

6 Final conclusion

The reported results substantiated that music training is an important aspect of human development and education, and may benefit all involved humans independently from any *congenital talent* or *gift*. Moreover, the results reported in the study have proved that music should not be abandoned and should be viewed as another factor that facilitates second-language acquisition.

Acknowledgments

The author would like to thank all participants for their kind participation in the study. Many thanks to Andrzej Wardak for the *ocen-wav* program development and Krzysztof Sztekmiler for his help with corpus preparation.

References

[1] J. S. Johnson, E.L. Newport, "Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language", *Cognitive Psychology* 21: 60-99 (1989)

[2] A.Baddeley, S.Gathercole, C.Papagno, "The phonological loop as a language learning device", *Psychological Review* 105 (1): 158-173 (1998)

[3] G.Bailly, "Close shadowing natural versus synthetic speech", *International Journal of Speech Technology* 6, Kluwer Academic Publishers, 11-19 (2003)

[4] D.Markham, "Phonetic, Imitation, Accent, and the Learner". Lund University Press (1997)

[5] B.Pastuszek-Lipińska, "The influence of music education and training on second language acquisition", Unpublished doctoral thesis, Adam Mickiewicz University (2007)

[6] B.Pastuszek-Lipińska, "Unpublished test of musical abilities" (2003)

[7] W.Sobkowiak, "English phonetics for Poles", Poznań: Wydawnictwo Poznańskie (2001)

[8] Ch.Dodane, "L 'apprentissage précoce d'une langue étrangère: une solution pour la maîtrise de l'intonation et de la prononciation?", in: E. Guimbretière (ed.), *La prosodie au coeur du débat: Apprendre, enseigner, acquérir*. Dyalang. Rouen: Presses Universitaires 312, 229-248 (2000)

[9] D.Birdsong, "Native like pronunciation among late learners of French as a second language", in: O.-S. Bohn and M.J. Munro (eds.), *Language experience in second language speech learning. In honor of James Emil Flege*. *Language Learning and Language Teaching* 17, Amsterdam: John Benjamins Publishing Company 99-116 (2007)

[10] M.Wrembel, "Phonological metacompetence in the acquisition of second language phonetics", Unpublished doctoral thesis, Adam Mickiewicz University (2005)

[11] L.Gilleece, "An empirical investigation of the association between musical aptitude and foreign language aptitude", Unpublished doctoral thesis, University of Dublin, Trinity College (2006)