Current State of the Development and Use of the "IntonTainer" System

1st Prof. Boris Lobanov
The United Institute of Informatics Problems
of National Academy of Sciences of Belarus
Minsk, Belarus
lobanov@newman.bas-net.by

2nd Vladimir Zhitko
The United Institute of Informatics Problems
of National Academy of Sciences of Belarus
Minsk, Belarus
zhitko.vladimir@gmail.com

Abstract—The purpose of this work is to describe a current state of the development and use of the "IntonTainer" system that provides a variety of ways to analyze and visualize the human intonation. The processing algorithms of a separate library, which became the core of the next version of the IntonTrainer application is described. Information on the use of the developed system is provided.

Keywords—Speech intonation, melodic portrait, intonation analysis, software model, IntonTrainer users

I. Introduction

The history of computer aided learning of pronunciation of a foreign learned language traces back at least to 1964, with the first attempt to display the learner voice pitch on a screen in order to improve the perception and the realization of sentence intonation [1]. Works in the last two decades show that the Computer-Assisted Pronunciation Teaching (CAPT) systems are useful, flexible tools for giving pronunciation instructions and evaluating at subject's speech [2]–[8]. However, pronunciation teaching has many issues. There are pedagogical, technological questions. From a technological viewpoint it is hard to provide understandable, accurate feedback to pronunciation mistakes. It is almost impossible to give a one hundred percent accurate and automatic diagnosis of speech production errors.

Earlier we proposed a software system for the analysis and training of speech intonation IntonTrainer [9], [10]. The system was designed to train learners in producing a variety of recurring intonation patterns of speech in foreign language. The system is based on comparing the melodic (tonal) portraits of a reference phrase and a phrase spoken by the learner. The main algorithms used in the training system proposed for analysing and comparing intonation features are considered. The proposed computer trainer provides additional visual feedback, as well as a quantitative estimation of the correctness of speech intonation in the process of foreign languages learning.

II. Current state of the development of the "IntonTainer"

This version provides many new possibilities for the analysis and training of intonation, namely:

- The ability to set the desired number of test files (pronunciations) and to obtain an average rating of their similarity with the reference phrase is added
- An auxiliary software module (Multi-Lingual Launcher) for launching any desired set of IntonTrainer modules is added
- The new IntonTrainer modules: Belorussian and Singing Lessons are added
- Added displaying the position of the boundaries of voice sections and the boundaries of accent units of phrases.
- The ability to analyze and compare the melodic portraits without the need for preliminary manual marking of it into pre-nuclear, nuclear and nuclear sections is added.
- The possibilities of parametric display of melodic intonation portraits are expanded.
- The ability to analyze and store in numerical form a set of prosodic signs of the analyzed phrases has been added.
- The accuracy of segmentation and labeling of the analyzed speech signals is increased.

"Fig. 1" show the view of the starting window of the Multi-Lingual Launcher program in the case when the user would need a full set of software modules. Using this service program, the user has an ability to promptly call one or more modules of interest to it.

The IntonTrainer application can be used not only as an intonation simulator, but also as an environment for conducting intonation studies. However, to date, its integration into other speech processing facilities has been difficult due to the lack of an external software interface for the IntonTrainer system. In this regard, it was decided to separate the processing algorithms into a separate library, which became the core of the next version of the IntonTrainer application. The structure of the core is shown in the diagram in "Fig. 2".

Figure 1. The starting window of the Multi-Lingual Launcher.
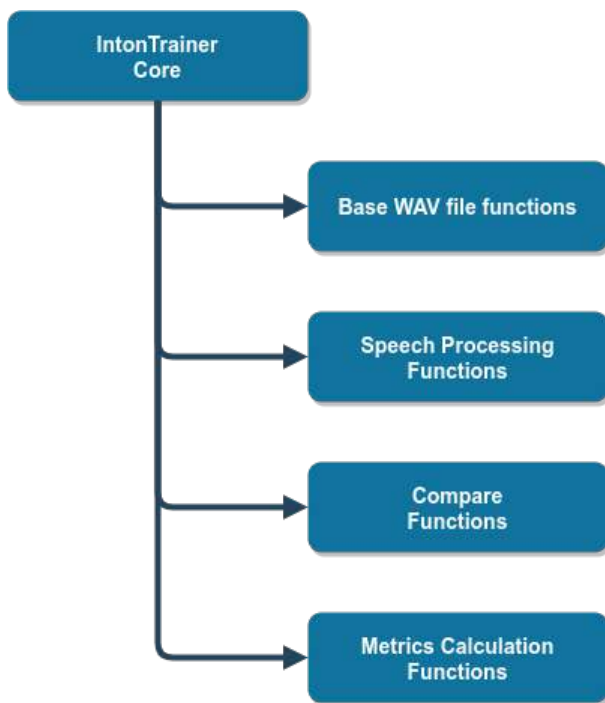

Figure 2. The structure of the system core.

The core includes the following main tasks for calculating speech flow parameters:

Pitch processing:

- get_pitch - the vector of the fundamental frequency (F0)
- get_pitch_log - logarithm vector from F0
- get_pitch_cutted - the frequency response vector taking into account the excess energy level
- get_pitch_min - minimum F0
- get_pitch_max - maximum F0
- get_pitch_interpolated – F0 vector with interpolated gaps
- get_pitch_norm - normalized vector of recovered F0
- get_pitch_smooth - smooth F0 vector

UMP processing:

- get_ump - UMP vector constructed by F0
- get_ump_mask - mask for mapping UMP to the original signal
- get_p - set of pre-nuclea UMP segments
- get_t - a set of segments of the post-nuclea UMP
- get_n - set of UMP nuclea segments

Signal energy processing:

- get_intensive - speech signal energy vector
- get_intensive_norm - normalized energy vector of a speech signal
- get_intensive_smooth - smoothed energy of a speech signal

Spectrum processing:

- get_spectrum - spectrum of the speech signal
- get_spectrum_norm - normalized spectrum of a speech signal
- get_cepstrum - speech signal cepstrum
- get_cepstrum_norm - normalized speech signal cepstrum

The calculation tasks are arranged in a certain hierarchy, where functions of a higher level use the results of calculations of the underlying functions. For each speech stream, the core creates a new object that caches the calculation results, which allows you to get different data on the same speech stream with less overhead, because Interim results will already be calculated.

The structure of the methods for processing speech flows is presented in the diagram of "Fig. 3".

Each method in the core can be configured using the core tuning mechanism. To store the settings, it is possible to download and upload to an external file. Also, one of the main parts of the core are methods for comparing two or more speech signals and calculating signal metrics and comparing them.

These methods include:

- Evaluation of the similarity of the ranges of variation of F0

  Range Similarity = (F0max / F0min) user * 100% / (F0max / F0min) template
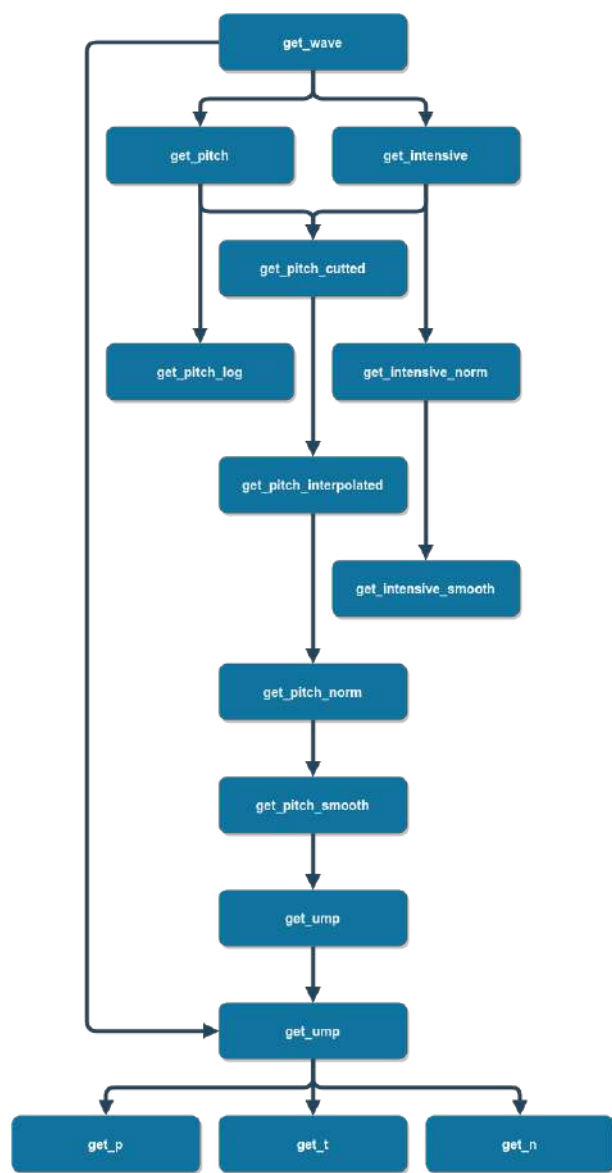
240

Figure 3. The structure of the methods for processing speech flows.

- Correlation assessment of similarity of UMP
  Shape Similarity (thru correlation) = [(r + 1) / 2] *
  100%
- The average value of the similarity assessment UMP
  Shape Similarity (thru average distance) = (1 - D)
  * 100%
- Minimum assessment of the similarity of UMP
  Shape Similarity (thru max local distance) = (1 - d
  max) * 100%

Because the core library is written in pure C, this will allow you to create wrappers for other programming languages, such as python, java. Creating wrappers for other languages can be automated using SWIG. Highlighting the main algorithms of the application will also allow IntonTrainer to be used as a plug-in in other research in automatic signal processing environments.

III. CURRENT STATE OF USE OF THE "INTONTAINER" SOFTWARE PACKAGE

The "IntonTainer" software package is recommended for use in the following popular fields:

- In linguistic education used as a means of visualizing intonation.
  (Primary introduction and study of the basic tone patterns of oral speech and songs, their pairwise comparisons, peculiarities of their usage as well as their actualization in dialogues prose and poetry).
- In individual intonation training for correct pronunciation used as a means of feedback.
  (Individual training for correct pronunciation of tone patterns when studying a foreign language, improving intonation skills of one's native language or singing in some professions, such as call center operators, radio and TV announcers, singers etc).
- In scientific and practical research used as a means of comparing intonation from different sources.
  (Study of individual, emotional and stylistic features of intonation. Comparative evaluation of speech intonation in norm and pathology. Estimation of the intonational quality of synthesized speech).

The "IntonTrainer" software package has been available for free download from the site https://intontrainer.by since the end of 2018. The world map ("Fig. 4") shows the geographical points, from which the requirements for the site for downloading the IntonTrainer came. The information in "Fig. 4" obtained using the site https://www.google.com/analytics/web/ and presented at the beginning of 2020.

Figure 4. The world map of the "IntonTrainer" downloading.

"Fig. 5" shows the distribution of downloads for the countries (except Belarus) with more than 5 download.

IV. CONCLUSION

The present prototype of the IntonTrainer system is implemented as a separate SWM under Windows OS (xp, 7, 8, 10) and Linux (Debian, Ubuntu). In order to make it possible to significantly increase the number of users,
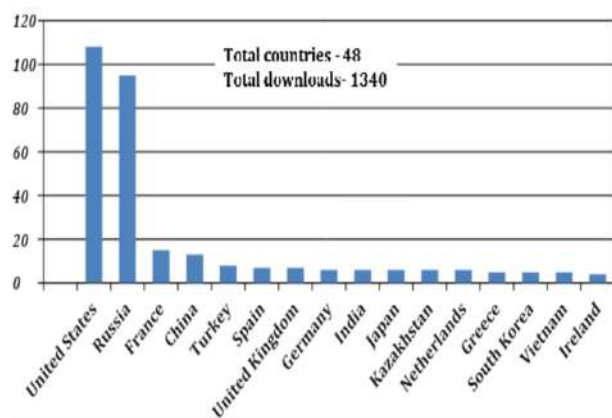
Figure 5. The distribution of downloads for the countries.

Современное состояние разработки и использованиясистемы «ИнтонТайнер»

Лобанов Б. М., Житко В. А.

Цель данной работы - описать текущее состояние разработки и использования системы «IntonTainer», которая предоставляет различные способы анализа и визуализации интонации человека. Описаны алгоритмы обработки отдельной библиотеки, ставшей ядром следующей версии приложения "IntonTrainer". Предоставляется информация об использовании разработанной системы.

we plan to develop an advanced version. This version should be easily portable to mobile and cloud platforms.

REFERENCES

[1] Vardanian, R. M. Teaching English through oscilloscope displays. Language Learning, 3/4, 109-118, 1964.

[2] Juan Pablo Arias a, Nestor Becerra Yoma a, Hiram Vivanco, "Automatic Intonation Assessment for Computer Aided Language Learning", Speech Communication, vol. 52, no. 3, pp. 254-267, Mar. 2010.

[3] Akinori Ito, Tomoaki Konno, Masashi Ito and Shozo Makino, "Evaluation of English Intonation based on Combination of Multiple Evaluation Scores", INTERSPEECH-2009, pp. 596-599. 2009.

[4] Motoyuki Suzuki, Tatsuki Konno, Akinori Ito and Shozo Makino, "Automatic Evaluation System of English Prosody for Japanese Learner's Speech", Proc. 5th Int. Conf. Education and Information Systems, Technologies and Applications (EISTA), 2007.

[5] Kim and W. Sung. "Implementation of intonational quality assessment system", INTERSPEECH-2002, pp. 1225–1228, Sept. 2002.

[6] Anthony Fox, "Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals", Oxford: Oxford University Press, 2000.

[7] Kazunori Imoto, Yasushi Tsubota, Antoine Raux, Tatsuya Kawahara, Masatake Dantsuji, "Modeling and automatic detection of English sentence stress for computer-assisted English prosody learning system", Acoustical science and technology 24(3), 159-160, May 2003.

[8] Konno, T. Ito, A. Ito, M. Makino, S. Suzuki, M. Grad. Sch. Of Eng., Tohoku Univ., Sendai, "Intonation evaluation of English utterances using synthesized speech for Computer-Assisted Language Learning", Int. Conf. Natural Language Processing and Knowledge Engineering, pp. 1-7, Oct. 2008.

[9] Lobanov, B. A Prototype of the Software System for Study, Training and Analysis of Speech Intonation / B. Lobanov, V. Zhitko, V. Zahariev // Speech and Computer: 20th International Conference, SPECOM 2018, Leipzig, Germany, September 18–22, 2018, Proceedings / – Springer, 2018. – P. 337-346.

[10] Lobanov B. Software Subsystem Analysis of Prosodic Signs of Emotional Intonation / B. Lobanov, V. Zhitko // Speech and Computer: 21st International Conference, SPECOM 2019, Istanbul, Turkey, August 20–25, 2019, Proceedings / – Springer, 2019. – P. 280-288.