

## **SRoL - Web-based Resources for Languages and Language Technology e-Learning**

S.M. Feraru, H.N. Teodorescu, M.D. Zbancioc

### **Silvia Monica Feraru**

Institute for Computer Science of the Romanian Academy, Iași, Romania  
E-mail: mferaru@etti.tuiasi.ro

### **Horia-Nicolai Teodorescu**

Institute for Computer Science of the Romanian Academy, Iași, and  
Technical University "Gheorghe Asachi" of Iași, Romania  
E-mail: hteodor@etti.tuiasi.ro

### **Marius Dan Zbancioc**

Institute for Computer Science of the Romanian Academy, Iași, and  
Technical University "Gheorghe Asachi" of Iași, Romania  
E-mail: zmarius@etti.tuiasi.ro

**Abstract:** The SRoL Web-based spoken language repository and tool collection includes thousands of voice recordings grouped on sections like "Basic sounds of the Romanian language", "Emotional voices", "Specific language processes", "Pathological voices", "Comparison of natural and synthetic speech", "Gnathophonics and gnathosonics". The recordings are annotated and documented according to proprietary methodology and protocols. Moreover, we included on the site extended documentation on the Romanian language, on speech technology, and on tools, produced by the SRoL team, for voice analysis. The resources are a part of the CLARIN European Network for Language Resources. The resources and tools are useful in virtual learning for phonetics of the Romanian language, speech technology, and medical subjects related to voice. We report on several applications in language learning and voice technology classes. Here, we emphasize the utilization of the SRoL resources in education for medicine and speech rehabilitation.

**Keywords:** spoken language resources, voice education, gnathosony, gnathophony, education, speech rehabilitation.

## **1 Introduction**

In a world where the Web and Internet communication is pervasive, the computer is more than a study topic for everyone, it is a ubiquitous tool. Computers serve for more than doing computations, they are now one of the most used means of communication and interaction - the very basis of any educational system. As a consequence, computer-based education is an obvious choice whenever a distance separates the learner and the learning person. In a general sense, computer-based education and virtual education based on Internet is today an undeniable fact of life in every academic campus [28], [29]. While computers and the network are the means, the spoken language represents the prevalent support of communication in the teaching-learning process. Hence, the natural need to address e-learning and virtual learning of languages, phonetics, voice pathology, and other aspects related to voice and spoken language.

In view of the above, we built during a timeframe of about five years a web site that offers the possibility of teaching and learning various aspects on the Romanian language, based on an annotated corpus freely accessible on the Internet. The corpus is complemented with in-depth phonetic and linguistic analyses, moreover with specific tools accessible by users from everywhere through the

web [16], [17], [18], [19], [25]. This instrument has a high level of dimensionality and aims to cover numerous aspects of the language that are not typical features in language corpora. This makes this "corpus-tool" an unique instrument of its kind existing today in the domain [22].

During the recent years, we developed an emotional speech database which can help in education and re-education of speech, in diagnosis and treatment, and in learning a language aided by computer; examples of related published results are [5], [16], [22].

Voice and language e-education is a topic addressed by many research and educational groups. Solomon [13] studied the possibilities and issues of learning with and about computers in schools or in other learning environments. The Eric Education Resources Page shows the importance of computer assisted education of speech and voice [24]. On the other side, web-based educational resources and training have received attention during the last decade. Ake Olofsson [10] offers a simple method of compensation for word decoding problems, by using a computer which pronounces the words which can not be read. Olofsson developed a program for the IBM-PC/AT and a Scandinavian multilingual text-to-speech unit that children can use to read a textfile on the monitor and request using a mouse the pronunciation of any word from that text [10].

The computer-assisted learning language software helps the interaction between student and computer by speech, by sound effects, by animation, and by video. On the other hand, the interaction is restricted typically to the mouse and keyboard. An active interaction, through spoken language enhances the educational computer-based tools [1]. In computer-assisted language learning, speech recognition offers the possibilities to have an active participation by oral reading and conversation. The CALL system reported in [1] includes recordings spelled by native speakers. The user has the possibility to compare the quality of her pronunciation with model recordings.

In another direction of research, Warschauer [23] observes the uses of online communications for language teaching. He determined that the interest in this domain grows day by day. He proposed a conceptual framework for understanding the role of the interaction assisted by computer [23]. Lundberg considers the computer a tool of remediation in the education of students with reading disabilities as dyslexic students which can benefit by computer training in correct reading and spelling the words [9].

A speech database is a collection of files with sounds, structured according to its own purpose. The SRoL resource (corpus) is located at the address ([www.etc.tuiasi.ro/sibm/romanian\\_spoken\\_language/index.htm](http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/index.htm)). The initiator conceived SRoL as an Internet-based "dictionary of sounds and words" for the Romanian language supplemented with specific manifestations of voice (including pathologies) and various tools. The SRoL database includes files with vowels, consonants, diphthongs, sentences with emotional states, linguistic particularities for the Romanian language, dialectal voices, and gnathosonic and gnathophonic sounds. It is the first Internet based annotated database of emotional speech for the Romanian language and contains more than 1500 recordings in different coding formats (.wav, .ogg, .txt, 22 kHz sampling rate, 24 bit or 16 bit precision). The phonetic recordings in SRoL, which refer to an annotated emotional speech corpus (database), are registered to ORDA.

## 2 The SRoL resources and the SRoL web site

The SRoL corpus evolved from a small research and educational speech database around 1995 (see Annex 1). It currently includes several sections, all freely available on the web. The main sections are:

- i) Standard pronunciation of vowels, diphthongs, words and short sentences in Romanian; the recordings in this section are appropriate for learning correct pronunciation in Romanian, moreover for statistical research on the Romanian phonetics;
- ii) Special syntactic constructs (linguistic peculiarities), like double subject and apposition; this section is research-oriented;
- iii) Emotional voices;

- iv) Analytic comparison between the synthetic and natural speech [27];
- v) Dialectal utterances;
- vi) A small archive of gnathosonic/gnathophonic sounds (included in the general "Archive of Sounds").

Beyond the main sections, the SRoL site includes an introductory section on the phonetics of the Romanian language, descriptions of the recording protocols and descriptions of the methodology, analysis tools (free software), extended research documentation, a video application, references, and a list of potentially useful links. The SRoL team developed instruments for signal processing regarding the extraction of patterns from voice signals, and the computing of the fundamental frequency (pitch) traces, respectively the traces of formants  $F_1$ ,  $F_2$ ,  $F_3$ . The site offers, beside executable programs, descriptions for each of these tools. Those descriptions are intended for a "general use", offering elementary explanations and relevant references for a better understanding [4], [22].

In this paper, we provide details about applications of the SRoL corpus, available to the address [http://www.etc.tuiasi.ro/sibm/romanian\\_spoken\\_language/index.htm](http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/index.htm).

### 3 SRoL as support for learning the Romanian language

One of the goals of the SRoL web site is to provide a free Romanian database for students and researchers, for linguists, for teachers, in view of teaching, learning and analysis the Romanian language sounds. The database includes the pronunciation corpus and related documentation. The database contains among others, sections with:

- recordings of syllables and words pronounced in various contexts, like accentuated word, interrogative sentences, exclamations, various emotions conveyed by the speaker, etc. This part of the database is aimed as a source for concatenative synthesizers and as benchmark for the voice recognition systems (isolated words), based on statistical models of language and speech, as [26];

- files of sounds, syllables and words pronounced by persons with various pathologies; this section may be useful in medical and phonological researches;

- files with professional voices ("perfect" pronunciations), as well as non-professional voices, the "voices of the people in the street". For the moment, we concentrate on voices from the Iași region (East Romania) and middle area of Moldova.

Learning and teaching languages require well documented audio-visual tools that exemplify and fully explain spelling for a large variety of voices and contextual and emotional states. While former methods, like tape recordings and audio disks have been helpful, the multimedia Internet-based tools offer tremendously increased capabilities. SRoL represents such a tool for the Romanian language. Not only it is the first for the Romanian language, but its multidimensionality makes it somewhat unique and novel in concept for language learning and teaching in general.

As an example of use, consider the case of a foreign student who wants to improve her Romanian pronunciation by comparing the prosody of her voice with the prosody of native speakers. The student utters a sentence (from those included in the site), then opens WASP<sup>TM</sup> or another similar tool and displays the energy and fundamental frequency in her voice. She then compares these prosodic features to the ones of native speakers and tries to improve her prosody until she produces correct prosodic patterns. Also, the student can compare formant values and try improving the formants of the vowels she pronounces.

This instrument is useful for learning to improve speech communication, moreover for human-computer speech interaction, for security, for medical applications, for video-games and interactive TV, for teachers, in the study of the Romanian language, etc.

## 4 Applications in medical education and re-education of speech

Application fields like language learning, professional voice education, and voice rehabilitation and re-education for medical conditions have different requirements, moreover are based on different methods. On the other side, education in medicine (ORL, phoniatics, dentistry) and in logopedy are other fields of potential applications of speech resources. Further, voice analysis for diagnosis is a domain that has seen significant progresses in recent decades. Voice education is needed whenever a voice pathology including some neurologic and psychiatric disorders, or pathology of the vocal tract occurs. Several groups have addressed the voice re-education topic [9], [10].

### 4.1 SRoL resources for minor voice pathology

Till now, we included in SRoL words pronounced by persons with minor pathologies, as trembling voice. We have demonstrated in our research that splitting the signal in frequency bands that correspond to the peaks of the  $F_0$ - $F_1$  formants and respectively to the peaks of  $F_2$ - $F_3$  formants helps improving the discrimination process in a significant way. The use of fractal dimensions in assessing the jitter or shimmer in voice produce mixed results [21]. Adding other fractal dimension, the rate of recognition of the tremor segments in voice improves, but it still low [21]. The voice pathology section of the database is useful in medical and phonological researches. Also for medical education use, the site comprises a gnathosonic and gnathophonic corpus.

### 4.2 SRoL resources for gnathosony

The gnathosonic analysis refers to the analysis of sounds produced during occlusion, due to the closing of the mandible over the maxillary at some stage in masticatory-like movements. Watt (cited in [7], [8], [11], [12]) has initiated the analysis of these sounds with application to diagnosis of the state of the stomato-gnathic apparatus during the 1960s and 1970s. The method has seen some interest, but it is not yet a current method in clinical practice.

The shape of the envelope of an occlusal sound is determined by the number of occlusal contacts and by the dynamics of the terminal part of the occlusion, namely by the dynamics of the sliding of the teeth, from the first contact until the equilibrium position in occlusion. A characterization of the waveform should take into account the need to correlate the sound with the medically relevant processes of contact and sliding. A limit in the occlusal sound analysis has been the complexity and the variability of shapes of the sound wave. The envelope of a single contact sound is characterized by the rise and fall times, value of the maximum, duration of the maximum, and total duration. The rise and fall curves follow exponentially laws, whose constants are of interest in the classification of the occlusal dynamics. For gnathosonic purposes, the sound signal  $s(t)$  generated by occlusion (teeth impact when closing the mouth like for mastication) and discretized as  $s[n]$  is first filtered by an elementary high pass, differential filter,  $s[n+1] \leftarrow s[n+1] - s[n]$ . Then, the signal is filtered with a nonlinear filter introduced in [14]. The filter first extracts the rough envelopes, averages them, applies to them median filters, sums the two resulting envelopes, and then apply to the sum an averaging filter [14]:

$$u_{inf}[n] = \min_{k=-6, \dots, 0} s[n+k], \quad u_{sup}[n] = \max_{k=-6, \dots, 0} s[n+k]$$

$$v_{inf}[n] = \frac{1}{7} \cdot \sum_{k=-6}^0 u_{inf}[n+k], \quad v_{sup}[n] = \frac{1}{7} \cdot \sum_{k=-6}^0 u_{sup}[n+k].$$

The next stage in the filtering is constituted by the median filtering on a moving window, as

$$z_{inf} = \text{median}_{k=-6, \dots, 0} \{v_{inf}[n+k]\}, \quad z_{sup} = \text{median}_{k=-6, \dots, 0} \{v_{sup}[n+k]\}.$$

and the two envelopes are summed – actually, summed in the sense

$$y[n] = [|z_{inf}| + z_{sup}] / 2;$$

$$e[n] = \frac{1}{p+1} \cdot \sum_{k=-p}^0 y[n+k].$$

We used a window of width 6 ( $p = 5$ ) for the last averaging. The widths of the windows in the above operations depend on the signal sampling frequency used in the recording process. The envelope of the signal is determined by taking the maximal respectively minimal value in a moving window, according to a procedure similar to the one explained for the filtering process. The envelope,  $e(t)$ , is itself low-pass filtered and then used for determining the occlusal sound parameters. The heuristic procedure applied to determine the duration of the occlusal sounds by forming "binary" impulses during the valid occlusal sound is:

if  $(e[n] > c_1)$  and  $B(e[n-14], \dots, e[n+14]) > c_2$  then  $h[n] = 0.1$ ,  
 else  $h[n] = 0$ ,

where  $B$  is a binary function (taking only 0 and 1 values) defined by

$$B = [\max(e[n-14], \dots, e[n-1]) > c_3] \& [\max(e[n+1], \dots, e[n+14]) > c_3].$$

The constants were chosen semi-empirically, as a function of the amplitude of the signal,  $c_{1..3} \sim A_s$  where  $A_s$  is the average amplitude of the signal after filtering (actually, we used the average amplitude of the sum of the envelopes), and the window width, 14, is determined by tests. We used the values  $c_1 = 0.001$ ,  $c_2 = 0.01$ ,  $c_3 = 0.005$ , which correspond to the average signal  $A = 0.027$ , determined as explained. For a normalized amplitude  $A$ ,  $A = 1$ , the constants are about  $c_1 = 0.05$ ,  $c_2 = 0.4$ ,  $c_3 = 0.1$ . The detection procedure can be further improved by reducing the false positives by imposing that the skewness of the impulse is larger than  $+0.5$ ; typical values for the skewness are larger than 0.7, showing that the rise of the impulse is significantly faster than the decreasing part.

## 5 Research support in gnathophonics and gnathosonics

In previous researches, we identified several ways the pathology of the stomato-gnathic system influences the speech:

i) The lack of the frontal dentition, namely of the upper teeth, may dramatically change the spectrum of the fricative consonants.

ii) The lack of the upper teeth may significantly modify the spectrum of the dento-alveolar sounds  $t$ ,  $d$ ,  $n$ , and  $l$ . (Notice that these sounds are rather alveolar in English, while in some other languages, like Spanish and Romanian, they may be dental. Therefore, the influence of the dentition on phonation is language-dependent.)

iii) The limited mobility and the pain in the temporo-mandibular joint (TMJ) impedes the production of fast transient vowels, especially in the diphthongs where the second vowel is pronounced with a largely opened mouth, like  $oa$ ,  $ea$ ,  $ua$ .

iv) The uncertainty in uttering due to a forcing in the TMJ, or to a poor neuro-muscular control may produce a tremor of the voice (fast amplitude changes, errors in the attacks, i.e. error in transitory regimes etc.).

v) The neurological pathology of the buccal cavity may impede on the accuracy of the pronunciation, including deficient starting of the words.

vi) Defective mobile prostheses may produce extra sounds, especially when the mouth is fast opened for pronunciation, moreover, it may produce clicks before the utterances.

vii) Prostheses of the upper teeth that do not provide for a physiological "V" shaped space between the teeth impede on the pronunciation of the fricatives, for example *f*.

viii) Especially the fricative consonants and the labial vowels are affected by the state of the dental furniture.

The *s* consonant uniformly occupies a large spectrum for a healthy dental apparatus, while it has a multi-band spectrum when the upper front teeth are missing or have deficiencies. The pronunciation of *s* and *v* may become close to that of *f*. For subjects with mobile prostheses, we noticed an uncertainty in the starting of the uttering.

The difference ratio in amplitude spectra is a parameter defined as:

$$\Delta S = \sum_k \frac{|S_1(f_k) - S_2(f_k)|}{S_1(f_k) + S_2(f_k)}$$

where  $f[k]$  is the  $k$ -th frequency in the FFT (Fast Fourier Transform) power spectrum of the two sounds and  $S_{1,2}$  are the average power spectra of the two sounds. For two similar sounds uttered by the same speaker, a difference larger than 50% means that the sounds are clearly distinguishable, while a difference smaller than 10% means that the sounds are indistinguishable. For example, if the average spectra for two sustained utterances of *f* and *v* have a  $\Delta S$  index of 40%, they will be distinguished by a listener, while if  $\Delta S = 15\%$ , they will be confused. We proposed the sustained consonant differential analysis as a method to further assess the impairment of speech production due to dentition. For this test, two similarly produced sounds are generated in a sustained mode and their spectra contrasted. For example, the sounds *f* and *v* are both at least partly fricative (*v* can be a semi-vowel, only partly fricative) that may be poorly produced due to imperfect dentition or neurological control. We conclude this section by stressing that gnathophonic testing should become a standard test for the dentist in the near future. The knowledge in the field is only emerging today, and fully developed, commercial tools are yet lacking, but the importance of the domain can not be refuted [7], [8], [11], [12]. The proposed tests are non-invasive, objective, and purely instrumental, hence their importance in the evaluation of the health state of the buccal system. These methods can easily be extended to remote, web-based diagnosis.

In figure 1, we exemplify a gnathophonic (a) and gnathosonic (b) recording sounds (for the speaker 19743m). In figure 1(a), we exemplified recordings of the Romanian words "*vata*", "*fata*", "*var*", intended to obviate similarities and differences in the pronunciations (Fourier spectra) of the consonants *f* and *v*, in the same context (beginning of the word, same *\_CVC* structure, with the same vowels and consonants, and *\_* denotes the beginning of the word). This is one of the specific choices of words proposed by the second author to determine when dentition defects produce confusion in the *f* – *v* uttered sounds. By analyzing such recordings available at SRoL, students can learn how to differentiate the normal and pathological states.

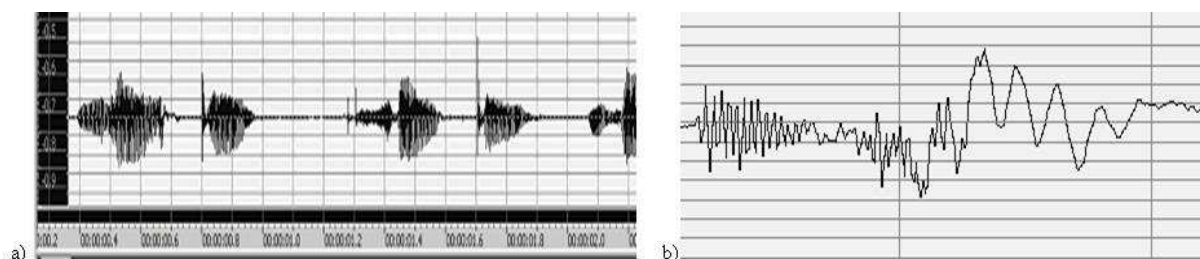


Figure 1: Gnathophonic (a) and gnathosonic (b) recording with details, tool GoldWave™

## 6 Applications in teaching the voice signal technology classes

Signal technology classes are taught around the world, especially for the master degrees in computer science and electrical engineering, moreover in some departments of linguistics and in a few medical centers. Some universities and education institutions developed their own databases and tools for speech processing. For examples, the Center for Spoken Language Understanding (CSLU) offers available language database from speech area and hearing science. These resources are important for analyzing the speech, for diagnosing and treating speech and language problems, for training students and so on. The tools and the corpora are distributed to over 2000 sites in 65 countries [2]. In education these tools help students learn about speech, learn a new language, learn through interactive media systems, or to become accustomed to hearing the normal and abnormal voice signal.

The second author currently uses the SRoL corpus in teaching and laboratory activities in the class "Speech Technology" given for the master degree in "Computational Linguistics" at the Faculty of Computer Science, "Al.I. Cuza" University of Iași. Details on the use in Voice Technology classes of some topics from SRoL are described in [4]. At the international EUROLAN 2007 summer school, the second author used the SRoL site to present "Traces of emotion, intentions and meaning in spoken Romanian" (<http://eurolan.info.uaic.ro/html/profs/HNTeodorescu.html>). The second author taught the specific methodology aspects, results obtained on the characterization of emotions in speech, possibilities of recognition of emotions and intentions in speech, and the relationship between specific meanings and the prosody in specific constructions in the Romanian language. The lesson exemplified applications of analysis of the speech emotional prosody to social, psycho-social, educational, and psycho-medical topics.

## 7 Software tools: pitch ( $F_0$ ) extractor

The extraction of the fundamental frequency  $F_0$  values combines four different methods: i) autocorrelation method (analysis in time domain) ii) the Average Magnitude Difference Function method, AMDF (analysis in time domain) iii) the Harmonic Product Spectrum method, HPS (based on spectral analysis) iv) the cepstral method (an analysis in que-frequency domain) - also applied for the higher formants searching.

The autocorrelation method is a classical method for pitch detection in the time domain. The method is based on the quasi-periodicity property of the voice signal and generates a local maximum that corresponds to the signal period. In the case of AMDF method, the local minimal values are detected and these values provide the necessary information to compute the fundamental period  $T_0$ .

$$C_k = \frac{1}{N} \cdot \sum_{n=0}^N x_n \cdot x_{n+k}, \quad k = \overline{0, W}$$

$$D_k = \frac{1}{N} \cdot \sum_{n=1}^N (x_n - x_{n+k}), \quad k \in \overline{1, W}$$

Here,  $C_k$  is the self-correlation,  $D_k$  is the difference function coefficient for a delay  $k$ ,  $x_n$  is the  $n$ -th sample of the signal,  $N$  is the number of correlation coefficients,  $W$  is the width of the analysis window.

The HPS method (Harmonic Product Spectrum) is based on the propriety that the spectrum of a periodic signal with fundamental frequency  $F_0$  has maximal spectral values at the multiples of this frequency  $2F_0, 3F_0, 4F_0, \dots$  (the harmonics of fundamental). When the signals are rescaled with the factors  $1/2, 1/3, 1/4, \dots$  after the decimation operation, by the multiplication of the resulted signals (which all have a spectral maximum in fundamental frequency  $F_0$ ), the other maximal value from spectrum are strongly attenuated.

$$H_n^k = H_{k \cdot n}^0 \text{ (decimation) or } H_n^k = \frac{1}{k} \sum_{i=0}^{k-1} H_{k \cdot n + i}^0$$

The cepstral method relies on the separation of the spectrum of the sound generator,  $H_g$  (which provide the information regarding the fundamental frequency), from the spectrum of the vocal signal filter,  $H_f$  (which describe the resonating cavities model). In the cepstral formula, the multiplication operation between the excitatory signal and the transfer function spectrums is transformed using logarithms into an addition operation:

$$\begin{aligned} H(\omega) &= FFT(s) = H_g(\omega) \cdot H_f(\omega) \\ cepstrum &= IFFT(\log|FFT(s)|^2) = IFFT(\log|H_g(\omega) \cdot H_f(\omega)|^2) \\ cepstrum &= IFFT(\log|H_g(\omega)|) + IFFT(\log|H_f(\omega)|^2) \end{aligned}$$

where  $FFT$  is the Fast Fourier Transform, and  $IFFT$  is the inverse  $FFT$ .

The results of the  $F_0$  extraction methods are compared in a decisional block, and a selection algorithm is used if there are significant differences. Another algorithm compares a current value with a number of neighboring values in order to select the nearest one, moreover compares the current values with mean values of  $F_0$ .

The error correction of the  $F_0$  extractors is performed through three methods:

- comparing the "neighbors": use the results provided by the same  $F_0$  extractor and if a difference between two consecutive values greater than a specified threshold value (usually 10-20%) is detected, the corresponding samples are considered errors;
- if the difference in absolute value between the current value of  $F_0$  and the average of fundamental frequency is greater than twice the standard deviation, then we consider those values as erroneous;
- if the current value of  $F_0$  is below 60% or over 150% of the average values of  $F_0$ , then we consider that the corresponding value is incorrect.

The threshold values were empirically determined and the final correction is accomplished by applying all the three correction methods described. The decision block receives the  $F_0$  values provided by the detection methods (AMDF differences method, autocorrelation method, HPS method, and cepstral method). To achieve the best possible pitch detection, the output values are weighted according on the performance of each  $F_0$  extractor. We assign smaller weights to the methods with a higher probability of providing incorrect outputs. The false detections of the fundamental frequency often consist in selecting the first subharmonic, or the first harmonic of  $F_0$ . When these "false" detection are not repaired by the correction module, we have two options:

- comparing the outputs of different  $F_0$  detection methods for the same window of analysis;
- comparing the outputs with a number of previous final results provided by the decision block.

## 8 Discussion

Our team has a long standing experience with using novel technologies in teaching, lasting for three decades [3], [7], [15], [20]. We applied that experience to the SRoL e-teaching and e-learning resource.

The SRoL resource is a vast annotated corpus of speech files complemented by tutorials, papers and additional files, moreover with tools for speech processing. If used by an experimented student or teacher, it may become a powerful tool for instruction and learning the Romanian language pronunciation, speech technology, and voice pathology and re-education. The SRoL sound voice resource is useful in many domains, including phonology, applied computer science, and medicine. Students and researchers may use this freely accessible site for learning the pronunciation of Romanian language, for



making comparative study between Romanian and other languages, for development of synthetic voice systems, for other linguistic, phonetic, socio-linguistic or medical applications.

This database is structured corresponding to precise criteria, documented and annotated according to a well defined methodology. The site has more than 1500 recordings of syllable, word, and sentence with various tonalities and pronounced with various emotional states. The database contains recordings of professional and normal voices, from the North-East region of Romania, without dialectal accent.

The SRoL resources have been recognized by several bodies, beyond the scientific publications that included our papers on SRoL. CLARIN European Network of Language Resources accepted SRoL as a member; ORDA (the Romanian Office for Authorship Rights) registered the original recordings, and the SRoL received a gold medal and media attention at the INVENTICA 2009 fair for inventions and creativity. Also, the website of Embassy of France in Romania briefly described in its Bulletin the SRoL site and its use in education (<http://www.bulletins-electroniques.com/actualites/58811.htm>). The Technical University "Gheorghe Asachi" of Iași intends to use SRoL in helping foreign students enrolled at this university to learn the correct Romanian pronunciation.

We hope the SRoL resources will be used in all the universities in Romania by foreign students who learn the Romanian language, moreover in other academic media and as an online tool by foreign students and teachers. We welcome any request for help and educational advice from all those who wish to use SRoL and the language-related web resources in virtual e-teaching and for e-learning.

## 9 Conclusions and future work

The SRoL speech annotated corpus constitutes the first extensive educational and research web speech corpus for the Romanian language. We believe it also constitutes a speech repository unique in many respects, including the first international language and sound resources for gnathophony and gnathosony, the first resources for comparative study of appositions and double subject constructions, moreover specific features as the rigorous methodology of documenting the records we used.

The objectives for the next two years are to increase the speech data base by about 1000 annotated recordings and to significantly extend the medical-oriented section of the resources. Also, we intend to add more tools for speech processing, including statistical tools on the GRID.

## Acknowledgements

The authors have been partly supported by the Romanian Academy, moreover the second author has been partly supported by a grant of the Ministry of Education and Science of Romania, during 2005-2006.

### NOTICES

1. A partial version [6] of this paper was presented in the ICVL 2009 conference and received the INTEL Special Award for Education (2009).

2. The authors contributions: the gnathophonic and gnathosonic research was been performed by the second author who also wrote the corresponding section of the paper (Sections 2, 4, 5, 6, and 8, and contributed to writing the other sections); the first author helped with further recordings and with their inclusion on the web page.

## Bibliography

- [1] K. Cameron, Computer Assisted Language Learning (CALL) Media, Design, and Applications, *Taylor & Francis*, ISBN: 902651543X, [http://www.google.com/books?id=dO\\_sNQIWhrsC & printsec=frontcover & dq=related, ISBN0940753030 & hl=ro & source=gbs\\_ similarbooks\\_s & cad=1](http://www.google.com/books?id=dO_sNQIWhrsC&printsec=frontcover&dq=related,ISBN0940753030&hl=ro&source=gbs_similarbooks_s&cad=1).
- [2] R.A. Cole, Tools for Research and Education in Speech Science, *Proc. Int. Conf. for Physics Students*, 1999, [www.cslu.ogi.edu/toolkit/pubs/pdf/cole\\_ICPS\\_99.pdf](http://www.cslu.ogi.edu/toolkit/pubs/pdf/cole_ICPS_99.pdf).
- [3] F. De Coulon, E. Forte, D. Mlynek, H.N. Teodorescu, St. Suceveanu, Subject State Analysis by Computer in CAE, *Proc. Int. Conf. on Intelligent Technologies in Human-Related Sciences*, Leon, Spain. Vol .2, pp. 243-250, 1996.
- [4] D. Cristea, H.N. Teodorescu, D.I. Tufis, Student Projects in Language and Speech Processing, *4th Conf. on Language Resources and Evaluation, Lisbon, Portugal Workshop on Language Resources: Integration and Development in E-learning and in Teaching Computational Linguistics*, pp. 17-22, 2004, <http://nats-www.informatik.uni-hamburg.de/view/Main/AcceptedPapers>.
- [5] M. Feraru, H.N. Teodorescu, The Emotional Speech Section of the Romanian Spoken Language Archive, *Conf. on Intelligent Systems and Technologies, Proc. 5th European*, Iași, Romania, ISBN 978973730497, 2008.
- [6] M.S. Feraru, H.N. Teodorescu, SRoL - Web-based Resources and Tools used for Language and Language Technology e-Learning, *Virtual Learning - Virtual Reality, Proc. 4th International Conference on Virtual Learning, ICVL 2009*, Bucharest University Press, ISSN: 1844-8933, Section Models & Methodologies, pp. 119-127, 2009.
- [7] W. Hedzelek, T. Hornowski, Gnathosonic Study of Occlusion in Patients Wearing Complete Dentures, *Eur J Prosthodont Restor Dent.*, Vol. 5, No. 3, pp. 119-23, 1997.
- [8] W. Hedzelek, T. Hornowski, The Analysis of Frequency of Occlusal Sounds in Patients with Periodontal Diseases and Gnathic Dysfunction, *J Oral Rehabil.*, Vol. 25, No. 2, pp. 139-45, 1998.
- [9] I. Lundberg, The Computer as a Tool of Remediation in the Education of Students with Reading Disabilities: A Theory-Based Approach, *Learning Disability Quarterly, Technology for Persons with Learning Disabilities*, Vol. 18, No. 2, pp. 89-99, 1995 <http://www.jstor.org/pss/1511197>.
- [10] A. Olofsson, Synthetic Speech and Computer Aided Reading for Reading Disabled Children, *Journal: Reading and Writing*, Vol. 4, No. 2, pp. 165-178, ISSN: 09224777, 1992 (<http://www.springerlink.com/content/j521536n135x2864/>).
- [11] J.F. Prinz, Computer Aided Gnathosonic Analysis: Distinguishing Between Single and Multiple Tooth Impact Sounds, *J Oral Rehabil.*, Vol. 27, No. 8, pp. 682-689, 2000.
- [12] J.F.Prinz, K.W. Ng, Characterization of Sounds Emanating from the Human Temporomandibular Joints, *Arch Oral Biol.* Vol. 41, No. 7, pp. 631-639, 1996.
- [13] C. Solomon, Computer Environments for Children - A Reflection of Theories of Learning and Education, 1988 [www.google.com/books?id=EonPZ9A81kkC&printsec= frontcover & hl=ro & source=gbs\\_v2\\_ summary\\_r& cad=0](http://www.google.com/books?id=EonPZ9A81kkC&printsec=frontcover&hl=ro&source=gbs_v2_summary_r&cad=0).
- [14] H.N. Teodorescu, Occlusal Sound Analysis Revisted, *Proc. 3rd Int. Conf. MEDSIP 2006, Advances in Medical, Signal and Information Processing*, ISBN: 0863416586, Glasgow, UK, 17-19 July 2006.

- [15] H.N. Teodorescu, Computer Semiotics: Understanding Meanings and Parallel Languages (Refereed invited paper) T. Yamakawa, G. Matsumoto (Eds.), *Proc. Int. Conf. IIZUKA'98, World Scientific Publ.*, pp. 279-283, 1998.
- [16] H.N. Teodorescu, M. Feraru, Classification in Gnathophonics - Preliminary Results, *The Second Symposium on Electrical and Electronics Engineering*, Galati University Press, pp. 525-530, ISBN 1842-8046, 2008.
- [17] H.N. Teodorescu, M. Feraru, Micro-corpus de Sunete Gnatosonice si Gnatofonice, Pistol, Cristea, Tufis (Eds.) *Resurse lingvistice si instrumente pentru prelucrarea limbii romane*, Ed. Universitatii "Al.I. Cuza" Iasi, ISBN 978-973-703-297-3, pp. 21-30, 2007.
- [18] H.N. Teodorescu, M. Feraru, D. Trandabat, Studies on the Prosody of the Romanian Language: The Emotional Prosody and the Prosody of Double-Subject Sentences, C. Burileanu, H-N. Teodorescu, (Eds.) *Advances in Spoken Language Technology*, The Publishing House of the Romanian Academy, Bucharest, Romania, ISBN 978-973-27-1516-1, pp. 171-182, 2007b.
- [19] H.N. Teodorescu, M. Zbancioc, E. Mihailescu, Speech Technology and Bio-Medical Engineering Teaching Based on the Web-A new Tool and Case Study, *Int. Conf. on Interactive Computed Aided Learning*, Villach, Austria, 2006.
- [20] H.N. Teodorescu, A. Kandel, B. Paschall, Teaching Modern Chapters in Automata Theory and Formal Languages, (abstract in booklet of the Symposium.) *Symp. 21 Century Teaching Technologies*, Univ. South Florida, Tampa, USA 2000.
- [21] H.N. Teodorescu, R. Ganea, M. Feraru, A. Burlui, Assesment of Voice Quality Based on Nonlinear Dynamic Analysis, *Proc. of The 15th Int. Conf. on Control Syst. & Computer Sci.*, Bucharest, Romania, pp. 536-542, ISBN 9738449898, 2005.
- [22] H.N. Teodorescu, D. Tandabat, M. Feraru, M. Zbancioc, R. Luca, A corpus of the Sounds in the Romanian Spoken Language for Language-Related Education In: C.P. Pascual (Ed.), *Revisiting Language Learning Resources*, Cambridge Scholars Pub. (CSP),UK, Ch. 6, ISBN 1847181562, pp. 73-89, 2007.
- [23] M. Warschaue, Computer-Mediated Collaborative Learning: Theory and Practice, *The Modern Language Journal*, Vol. 81, No. 4, Special Issue - Interaction, Collaboration, and Cooperation - Learning Languages and Preparing Language Teachers (Winter, 1997), pp. 470-481, <http://www.jstor.org/pss/328890>.
- [24] B.W. Wise, R.K. Olson, Computer Speech and the Remediation of Reading and Spelling Problems, *J. Special Education Technology*, Vol. 12, No. 3, pp. 207-220, 1994.
- [25] M. Zbancioc, Tools for the Archive of the Romanian Language Sounds Project, *4th European Conf. on Intelligent Systems and Technologies*, Iasi, Romania, ISBN 973-730-265-6, 2006.
- [26] Kenko Ota, Emmanuel Dulfos, Philippe Vanheeghe, Masuzo Yanagida, Bayesian Inference for Speech Density Estimation by the Dirichlet Process Mixture, , *Studies in Informatics and Control Journal*, Bucharest, Romania, ISSN 1220-1776, Vol. 16, No. 3, 2007.
- [27] Florin Grigoras, Horia-Nicolai Teodorescu, Vasile Apopei, Nonlinear Analysis and Synthesis Of Speech, *Studies in Informatics and Control Journal*, Bucharest, Romania, ISSN 1220-1776, Vol. 7, No. 1, 1998.

- [28] Tom Page, Gisli Thorsteinsson, Andrei Niculescu, Management of Knowledge in a Problem Based Learning Environment, *Studies in Informatics and Control Journal - With Emphasis on Useful Applications of Advanced Technology*, Bucharest, Romania, Vol. 18, No. 1, 2009.
- [29] Antonios Andreatos, International Journal of Computers, Virtual Communities and their Importance for Informal Learning Communications and Control, *International Journal of Computers, Communications and Control - IJCCC*, Romania, ISSN 1841-9836, Vol. II, No. 1, pp.39-47, 2007.

### Annex 1. Development stages of SRoL

The presently named SRoL corpus started around 1995 as a small, research and educational database including examples of recordings with vowels and a few typical words in Romanian, moreover a few recordings of pathological voices. It was correlated to the class of Image and Speech Processing given by the second author in the "Gheorghe Asachi" Technical University of Iași, Romania. Former students (who are now professors in several Romanian universities) contributed to that incipient voice database (credit for recordings and other help for that database deserve the now professors Radu Ciorap and Irinel Pletea, among others). The database was further developed for educational purposes in relation to the the class of Speech Technology given by the second author in the Faculty of Computer Science of "Al. I. Cuza" University in Iași.

The third stage of development started in 2004, when the second author decided to significantly enlarge and move the speech database on the web, partly with the help of two grants that helped forming a team in the Institute for Computer Science of the Romanian Academy and in the "Gheorghe Asachi" Technical University of Iași. The first author joined the team, at that time as a fresh Ph.D. Student. Since the second author initiated five years ago the Project "The Sounds of Romanian Language" (SRoL), the team increased to 8 researches. The SRoL Web-based spoken language repository and tool collection as it is today was developed during several years by the collaboration of groups from the Institute for Computer Science of the Romanian Academy, CERFS Excellence Center in "Gheorghe Asachi" Technical University of Iași and by staff of the discipline of Language Technology, Computer Science Faculty, "Al.I. Cuza" University.

### Annex 2. Typical shapes of gnathosonic signals

The sketches below stand for the envelopes of typical gnathosonic signals, corresponding to normal, merged double contact, and isolated double contact signals. The sound is easily categorized by automatic means.

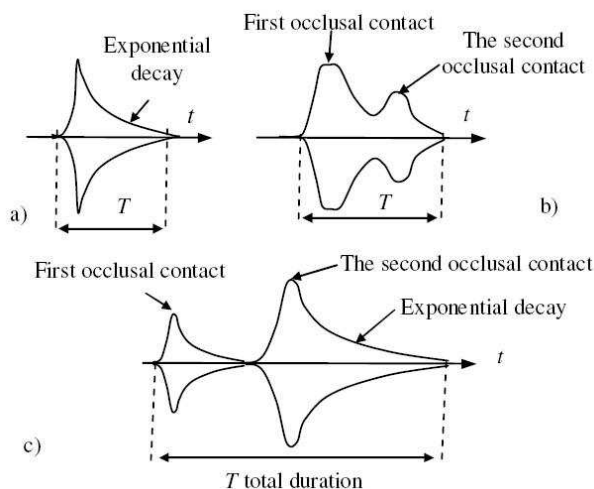


Figure 2: Typical envelopes of occlusal sounds (from [14])

**Silvia Monica Feraru** (November 21, 1977) received a MSc. degree in BioMedical Engineering (2004) and PhD in Electronics (2009) from "Gheorghe Asachi" Technical University of Iași. Now she is research assistant at the Institute for Computer Science of the Romanian Academy, Iași branch. She received the Special Awards Intel Education 2009 at The International Conference on Virtual Learning, ICVL 2009. Her current research interests include vocal signal processing, cognitive processes, and various aspects of artificial intelligence. She has (co-)authored more than 21 conference, journal or bookchapter papers.

**Horia-Nicolai Teodorescu** (November 14, 1951). MS in Electronics, "POLITEHNICA" University, Bucharest, 1975, Ph.D. in Applied Physics - Electronics, under the supervision of the late Prof. Emil Luca, at the Technical University of Iași, 1981. Currently, he is a professor at the "Gheorghe Asachi" Technical University of Iași and the director of the Institute for Computer Science of the Romanian Academy, Iași. He is a correspondent member of the Romanian Academy. Has authored or co-authored about 300 journal and conference papers, holds 24 national and international patents and has received numerous national and international awards and prizes. He is a Senior Member, IEEE.

**Marius-Dan Zbancioc** (August 15, 1975) teaching assistant at the "Gheorghe Asachi" Technical University of Iași and researcher at the Institute of Computer Science of the Romanian Academy, Iași branch. His current research interests include signal processing, expert systems, fuzzy systems and several aspects of artificial intelligence. He has (co-)authored 3 books and 39 papers.