STUDIA UNIVERSITATIS BABEŞ-BOLYAI, PSYCHOLOGIA-PAEDAGOGIA, LV, 1, 2010

VISUALIZING SPEECH ACOUSTIC FEATURES BY USING COMPUTERIZED PROGRAMS IN THE CONTEXT OF PROFOUND HEARING IMPAIRMENTS

CAROLINA HATEGAN

ABSTRACT. Through this article it is underlined the way two computerized programs can be complementary used (Cool Edit Pro and Praat) in order to visualize speech acoustic features in the context of profound hearing impairment. The linguistic material that was used was both isolated sounds and co-articulated sounds at word level. The participants in the research are profoundly hearing impaired persons, both using verbal language, one person wearing digital hearing prosthesis and one without benefits from technical aid.

Keywords: *hearing impairment, spectrograms, hearing aid, auditory-verbal therapy, vowels, consonants, speech spectrum*

ABSTRAKT. Durch dieser Arbeit wird die Art und Weise gezeigt, wie man ergänzend zwei Computerprogrammen (Cool Edit Pro und Praat) benutzen kann, um die akustischen Besonderheiten der Sprache zu visualisieren, bei einer tieferen Hörschwäche. Das benutzte Material sind isolierte Laute und koartikulierte Laute in einem Wort. Die Teilnehmer bei der Forschung sind zwei Personen mit tieferer Hörschwäche, es wurde ihnen die Sprache gelernt, eine mit digitale Prothese und eine ist ohne.

Stichwörter*: Hörschwäche, Spektogrammen, Hörprothesen, hör-mündliche Therapie, Selbstlaute, Konsonanten, bisyllabische Wörter.*

1. Theoretical back-ground

In order to facilitate the understanding of the aspects concerning speech analysis and synthesis, from a physical point of view a complete description of the fonator phenomenon is required (Trial, 1991).

Human phonation system is composed by three distinct components:

-lungs;

-larynx;

-supra-glottis cavities (pharynx, oral cavity and nasal cavity).

In the central area of the larynx are vocal cords located, these being either close related, or at distance leaving a triangular openness, o variable dimensions.

This openness is called glottis. Under the action of the impulses the muscles of the vocal cords receive from the recurrent nerves, they begin vibrating; the sound that goes along the glottis is transformed in larynx cvasi-periodical impulses, with wave form. Under the circumstances when this vibration has a non-sinusoidal form, it can be represented through an infinite sum of components: amplitude; frequency; time (phase). The decomposition in those three components is Fourier serial type.

The first harmonic to which the lowest frequency corresponds is called voice fundamental (F0), the other harmonics have frequencies equal to multiple integers of fundamental frequency (Coleman, 2005).

Human voice fundamental frequency varies between 60-70 Hz range, corresponding to the gravest male voices; till 1200-1400 Hz in sopranos case, this range being the superior limit for voice fundamental. The media is between 100-150 Hz, for males; between 200-300 Hz for females and 350-500 Hz for children. While speaking, due to intonation, fundamental frequency varies in a range approximately of one octave (the octave is an interval of two frequencies, their report being equal with two) (Rosetti and Lazaroiu, 1982).

Graphical representations of the all harmonics amplitudes according to frequency is called spectrum. For cvasi-periodical larynx impulses that are characterized by a wave form, the spectrum varies inverse proportionally with the square route of the frequency.

Laryngeal impulses can’t be directly perceived because these are not in acoustic relation with the exterior but through supra-glottis cavities. Thus, supra-glottis cavities become a complex resonator, with features of variable filter that implies for laryngeal cvasi-periodical impulses important changes that consist in selective amplification of the harmonics placed on the frequencies that belong to the resonator chambers from the supra-glottis level.

The harmonics that were amplified through the resonator phenomenon by supra-glottis cavities are called formants. The formant is defined as a frequency area where the higher quantity of acoustic energy is concentrated. Formants distribution along the spectrum determines the tone of the perceived sound.

Linguistic system for any historical language is composed from the vowel and consonant system, having into consideration a certain linguistic perspective (Stan, 1996). This type of approach must be completed by including stress and intonation as equally important in defining a linguistic system (Anca and Hategan, 2008). In this article we will underline this aspect by emphasizing the way emotional tone influence the spectral representation.

2. Research objectives

- comparing spectral representations elaborated by two different software programs

- emphasizing articulator and co-articulator characteristics in the context of profound hearing impairment
- emphasizing differences in articulator and co-articulator abilities between profound hearing impaired persons with or without hearing aids.

3. Research hypotheses

- since the two programs were built on the same principle, they would not offer spectral contradictory data, but complementary one
- speech disorders (distortions, sound omissions) as well as voice disorders (head voice, hoarse voice, nasal voice) are present in articulator, as well as in co-articulator contexts, in the cases of both participants
- articulator and co-articulator abilities are superior in the case of the participant with hearing prosthesis, due to enhanced laryngeal control (controlled insured through auditory feedback)

4. Participants

Two participants were included in the research:

- one person with profound hearing impairment, adult, using verbal language, without wearing hearing prosthesis
- one person with profound hearing impairment, using verbal language, with digital prosthesis (with a significant hearing benefit from technical compensation)

5. Means and materials:

Selected instruments include: technical means and verbal material.

5.1 Technical devices

Two software programs were utilized, which allow for visualization of acoustic characteristics of speech. The programs are Cool Edit Pro and Praat.

5.1.1 COOL Edit Pro software

COOL Edit Pro software insures ways of recording, visualizing, and processing acoustic information. Program options include also filtering for environmental noise control. Recordings can be done on a broad frequency range, up to 10kHz, to cover for all the field of frequencies for speech. Image resolution is good enough to generate intuitive representations, easy to read and interpret. Through this program, one can visualize the shape of speech wave, and its spectral representation. The most significant visualizing modality is through the spectrogram (a three-dimensional graphical representation comprised of the following variables: time, frequency, intensity). Time is represented on the horizontal axis, in milliseconds. Frequency is represented on the vertical axis, in Hertz. Intensity can be visualized through chromatic representation.

In the case of vowels, in order to clearly visualize the first three formants, one can utilize a frequency range of 6 kHz. These formants are the indicators that differentiate among vowels. Formants are dependent on the way the sound is produced in larynges, and then filtered by speech organs, which modify the size and shape of resonator cavities.

Consonants can be recorded on a frequency range of 10kHz. In the case of consonants, their visualization is more difficult due to their dynamic character and dependency on associated vowels. Even so, spectrograms offer important data that allows for consonant discrimination, based on acoustic energy focus (voice bar for voice consonants, vibrations for “r” etc.).

To analyze co-articulator phenomenon, the program allows for spectrograms for diphthongs, mono, bi, and three syllabic words.

Another function of the program allows for graphical representation for analysis of frequency and intensity. These graphic representations allow for a more clear visualization of formants (amplified frequencies in the case of vowels) and noise. This analysis is more adjusted for vowels that are time consistent, when pronounced individually. For consonants, which are dynamic, frequency analysis would give a great number of graphic representations for each of their production steps.

With regards to intensity, one can emphasize its relative character. What remains constant is the intensity ratio between various amplified regions on frequency intervals, the ratio between the formants peaks in the case of vowels.

5.1.2 Praat software

PRAAT is a computerized program with recording, analysis, synthesis, speech sound manipulation functions, by means of high quality images.

The recorded material can be visualized in various ways, as frequency representation, spectral representation, audiogram etc. Spectral representation does not allow for chromatic coding of spectrogram, as in Cool Edit Pro program, but for chromatic emphasis pf pulses and formants. Consequently, the spectrogram is chromatically represented in grey and black, with areas of more intense grey to black nuances indicating a greater concentration of energy. The frequency range for both mono and stereo recording of sounds is between 0- 96000Hz, with an optimum value in 0-500 Hz interval.

5.2 Verbal material

Verbal material is comprised of vowel and consonant sounds. These are : a,i,o,u,ă,î, b,d,c,g,f,v,h,j,l,r,m,n,p,s,ş,ţ,z,t. To also emphasize minimum co- articulator abilities, participants were asked to produce also bi- phonemic structure “ca”. Also, co-articulator abilities were measured through the production of the bi-syllabic word “progres” (in English “progress”). The selection of this word was done on the phonemic criterion, this means that it does not imply complex co-articulator contexts, aspect that ensures the possibility to emphasize voice troubles. It is also aimed the correct co-articulated pronunciation of a fricative and of an explosive sound, but also of the vibrant “r”. Thus, this word offers the possibility to emphasize specific pronunciation impairments in the case of hearing impaired persons.

6. Results

By superposing the spectral representations it can be underlined the fact that those two programs ensure very resembling spectral representations, in the case of the isolated articulated sounds. In order to underline this aspect it can be illustrated through figures 1 and 2, figures that contain the edited spectrograms by using Cool Edit Pro and Praat soft- wares, for the vowels and consonants pronounced by the two participants in the research.

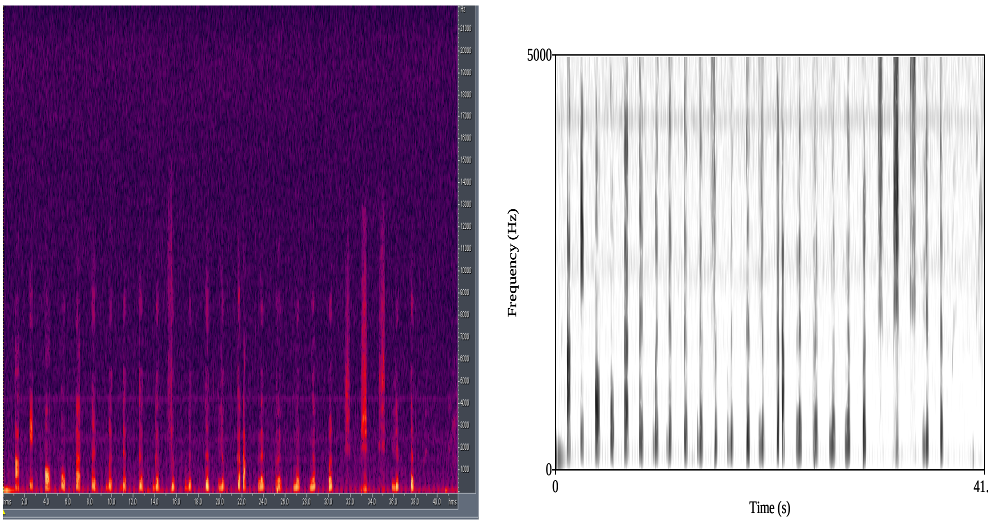


Fig. 1.

a **b.**

Spectrograms for vowels and consonants articulated by the participant with prosthesis
a. spectrogram edited with Cool Edit Pro program
b. spectrogram edited with Praat program

It is interesting to notice the fact that in the case of the isolated articulated sound, Cool Edit Pro program proves to be more efficient as the colored representation ensures a better possibility to perceive a reduced temporal extension spectral representation. In the case of spectral representations for co-articulated structures, Praat program utility can be underlined, thus those two programs can be complementary used in speech therapy field.

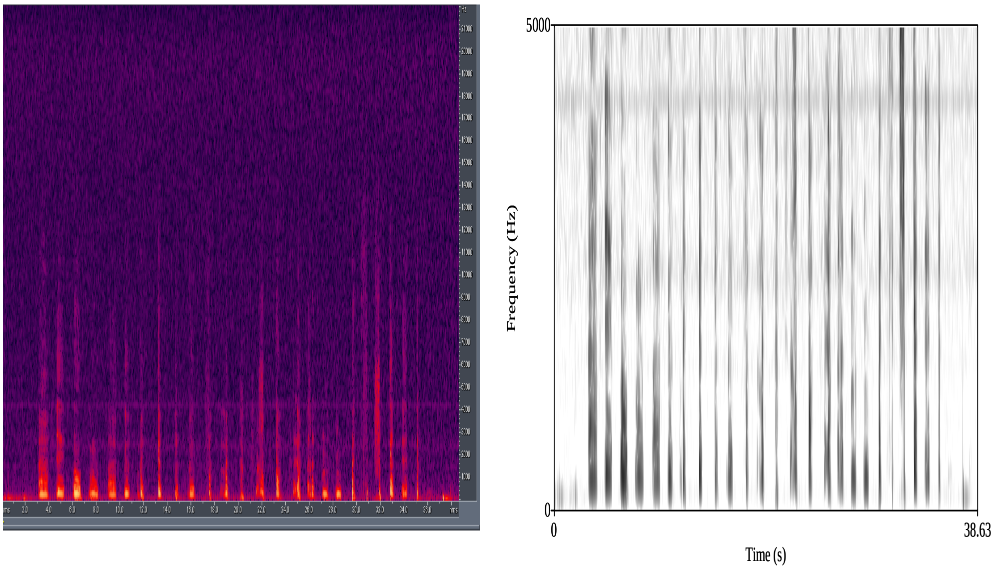


Fig. 2.

a. **b.**

Spectrograms for vowels and consonants articulated by the participant without prosthesis
a. spectrogram edited with Cool Edit Pro program
b. spectrogram edited with Praat program

If Cool Edit Pro program does not ensure the possibility to establish the frequency area, it can be observed the fact that spectral representation is less clear and the capacity to delimitate the formants from the background noise is diminished. By underlining the contributions of those two computerized programs for visualizing speech acoustic features in the case of hearing impaired persons, both concerning isolated sounds articulation, as well as co-articulated sounds (the pronunciation of complex linguistic structures) ensures a detailed perspective related to auditory-verbal therapeutically approach. Thus, the spectrograms for the word “progress” underline the fact that the most important difficulties concerning the way expressive language level is configures in the case of hearing impaired persons are directly dependent on co-articulating context, and not on the context of isolated sounds pronunciation. This aspect is also mentioned in the classical literature from this field, the computerized technique that ensures spectral visualizing just proving again an empirical aspect (Anca, 2007; Stanica et al. 1983, Pufan, 1972).

Analyzing the spectral representations for the vowels and consonants articulated by those two participants in the research, as well as the spectral representations for “progress” word, in the case on the same participants in the research, it can be noticed the fact that significant differences between at the level of their phonological abilities can be better emphasized by referring to the bi-syllable structure.

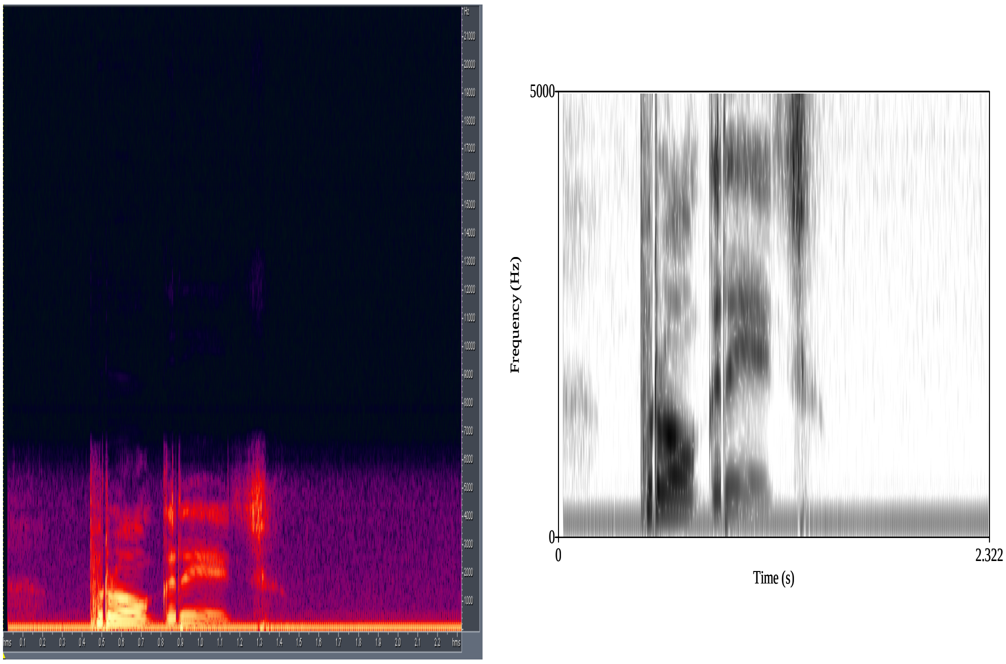


Fig. 3.

a. **b.**

Spectrograms for „progress” words articulated by the participant with prosthesis
a. spectrogram edited with Cool Edit Pro program
b. spectrogram edited with Praat program

Referring to the pronunciation of the word “progress” visualized with the help of Praat computerized program, in the case of profoundly hearing impaired person, with digital bilateral prosthesis the fallowing aspects can be mentioned:

- the formants can be visualized between the 1356 Hz and 4410 Hz range fallowing the pattern: F1=1356 Hz; F2=1607 Hz; F3=2997 Hz; F4=4410 Hz;
- voice fundamental can be delimitated at F0=458 Hz;
- pronunciation intensity is 67.216 db;
- pronunciation time is 1.16 s.

Pronunciation troubles in the case of this participant in the research consist in the presence of glottal stop of the first syllable is pronounced, but also after “gre” structure, “s” sound being isolated articulated, without co-articulation phenomenon to be noticed. The same aspect can be underlined at the level of the spectrogram with vowels and consonants (figure 1 and 2) for the articulation of the bi-phonemic construction “ca”, both in the case of the participant in the research with prosthesis, and in the case of the participant without prosthesis.

At the level of “progress” word the fricative phoneme “s” is pronounced with intensity, its articulation being marked by the presence of significant back-ground noise on higher frequencies. The fricative phoneme “s” is not pronounced as a voiced sound, aspect that indicates a functional control from the point of view of larynx activity, control due to the auditory feed-back (aspect that proves that the participant benefits from the technical compensation).

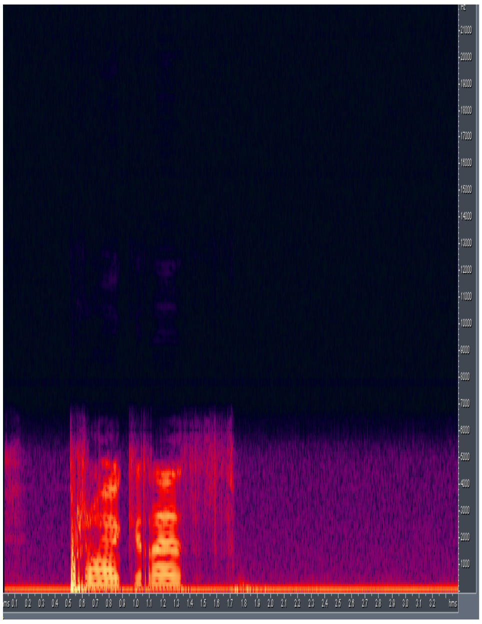
There are not to mentioned deficitar aspects concerning the way the explosive and the vibrant sounds were pronounced, aspect that underlines superior articulator and co-articulator abilities. On the contrary, there can be put in evidence voice troubles such as nasalization; troubles easy to be underlined with the help of Praat computerized program, due to the presence of white noise over the spectral representation. It is also important to notice the fact that the formants for “ e” and “o” vowels are to be put in evidence at the level of a little bit higher frequencies than those mentioned in the specialized literature as being specific for speech neutral tone (the first three formants for “o” vowel have the fallowing values: F1= under 500Hz; F2=1000 Hz; F3=2500 Hz, and for “e” vowel: F1=500 Hz; F2=2000 Hz; F3=2600 Hz; Anca and Hategan, 2008), In this way it can be underlined a weak control over the suprasegmental component of language and not the presence of voice troubles such as head voice. This aspect can be sustained by the high level intensity of the word pronunciation, 67, 216 db, it being also underlined the fact that the word is over-stressed, the linguistic structure being emphatically articulated (www.etc.tuiasi.ro/.../romanian.../index_nou.htm).

The pronunciation time is more limited than in the case of the person without hearing aids, 1,16 s, comparatively with 1,69 s, aspect that highlights the fact that co-articulator abilities are superior in the case of the hearing impaired person with digital prosthesis, due to a higher degree of laryngeal control.

Referring to the pronunciation of the word “progress” visualized with the help of Praat program, in the case of profound hearing impaired person without prosthesis, the fallowing aspects can be underlined:

- the formants can be visualized between the 787 Hz and 3731 Hz range fallowing the pattern: F1= 787 Hz; F2=1192 Hz; F3=2651 Hz; F4=3731 Hz;
- voice fundamental can be delimitated at F0=460 Hz;
- pronunciation intensity is 62.144 db;
- pronunciation time is 1.69 s.

The pronunciation spectrum is blurred and fragmentized, in the case of the profoundly hearing impaired person without prosthesis. This aspect is due to a deficitar control over the supra-glottis cavities, the resonance chambers being inadequately involved in co-articulating act. Thus, the formants are very difficult to be put in evidence, the back-ground noise being significantly amplified, aspect that confer a blurred character to speech spectrum. Nasal resonance, on low frequencies (in conferring a character of hoarse voice) also contributes to the spectrum fragmentation (Anca, 2007).



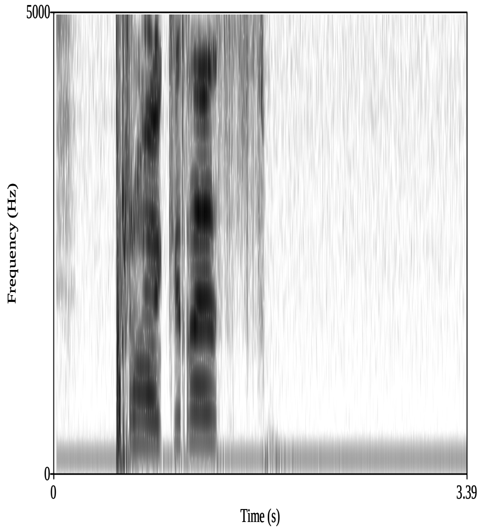


Fig. 4.

a. **b.**

Spectrograms for „progress" word articulated by the participant in the research without hearing prosthesis a. spectrogram edited with Cool Edit Pro program
b. spectrogram edited with Praat program

Despite de grave tonality of the voice, aspect underlined by the formants frequencies, in the case of this participant in the research can be emphasize the head voice, especially at the level of the spectrum corresponding to vowel "o", vowel that is articulated similar with vowel "e" (aspect that can be found also at the level of figure 2 and at the level of fricative sound "s" spectrum area).

As in the case of the participant in the research with prosthesis, in the case of this participant can be underlined the presence of the glottal stop after the first articulated syllable "pro", but also after the structure "gre". Thus, the fricative sound is isolated pronounced, the constrictive character being highly emphasized, especially on higher frequencies. At the level of this sound is to be highlighted the mixture between head and horse voice and the generalization of nasalization.

In this participant the laryngeal control is weaker in the absence of the auditory feed-back, aspect that confer a non-linear feature for the voice fundamental a-periodical oscillations being present at this level. This weak control of laryngeal activities in the case of hearing impaired person without hearing aid can be put in evidence by comparing figure 1 and 2, aspect also emphasized by specialized literature (Anca, 2007; Lewis, 1996; Manolache, 1980).

Conclusions

Analyzing the speech spectrograms for the two participants in the research, it is clearly emphasized the fact that the most numerous voice and pronunciation troubles both in the articulator level and in the co-articulator one are to be identified in the case of the person with profound hearing impairment without auditory prosthesis.

Thus, if the computerized program Praat allows to accurately calculate the coordinates on which speech can be visualized (frequency, intensity, time, formants), Cool Edit Pro program allows editing the sound by reducing back-ground noises and voice troubles such as nasalization (there can be applied filters for cleaning the spectrum). In this way there can be underlined the complementary utility of those two programs, especially in the context of hearing loss and verbal- auditory therapy, adult person with hearing impairment having the possibility to increase articulator and co-articulator control by using them, aspect also underlined by researches developed in the Anglophone area (Ertmer, 2004).

Another research direction that can be opened is the one of selecting linguistic material for auditory-verbal therapy or for the vocal audio-metrical measurements technique (for this audiometric procedure the linguistic material has not been selected in the psycho-pedagogical Romanian field since 1964 year, when it was selected by Constantinescu F.) being also used the computerized programs put in work within this research.

Acknowledgements

This work was supported by CNCSIS- UEFISCSU, project number PN-II IDEI 2449/2008.

BIBLIOGRAPHY

Anca, M. (2000). Examinarea şi evaluarea funcției auditive, Presa Universitară Clujeană, Cluj- Napoca.

Anca, M. (2007). Psihologia deficienților de auz, reeditata Presa Universitară Clujeană, Cluj- Napoca.

Anca, M.; Hategan, C. (2008). Terapia limbajului. O abordare interdisciplinara, Presa Universitara Clujeana, Cluj-Napoca.

Coleman, J. (2005). Introducing Speech and Language Processing, Cambridge University Press, Cambridge.

Constantinescu, F. (1964). Contribuții la fundamentarea științifică a audiometriei vocale. Probleme de defectologie, Editura Didactica si Pedagogica, Bucureşti.

Ertmer, D.J. (2004). How well can children recognize speech features in spectrograms? Comparisons by age and hearing status, Journal of Speech, Language and Hearing Research, indexed for Medline, 47 (3):484-95.

Lewis, S. (1996). Language Acquisition in Hearing-Impaired Children, University of Birmingham Press, Birmingham.

Manolache, C. (1980). Surdomutitatea. Ortofonie. Lmibaj. Auzitori şi surdomuți, Bucureşti

Pufan, C. (1972). Probleme de surdo-psihologie, Editura Didactica si Pedagogica, Bucuresti.

Rosetti, Al.; Lazaroiu, A. (1982). Introducere în fonetică, Editura Ştiințifică şi Enciclopedică, Bucureşti.

Stan, I.T. (1996). Studii de fonetică şi fonologie, Presa Universitară Clujeană, Cluj-Napoca.

Truxal, J. G. (1991). The age of electronic messages, MIT Press, Massachussetts.

http://www.fon.hum.uva.nl/praat/

www.etc.tuiasi.ro/.../romanian.../index_nou.htm