

MAŁGORZATA WARYSZAK
Maria Curie-Skłodowska University in Lublin
Department of Logopedics and Applied Linguistics

Acoustic analysis of normative and impaired emotional prosody*

SUMMARY

The present article describes how biologically normal persons and those with various types of speech disorders use the suprasegmental structures of the language in order to express emotions. It presents the results of pilot studies, based on an experiment. In order to objectify them, the method of speech acoustics was applied. The author seeks to characterize the expression of selected emotions using acoustic parameters of the speech signal in order to determine its universal patterns. She also draws conclusions about the impact of perceptual skills deficits (in hearing loss) and of articulation efficiency (in dysarthria in infantile cerebral palsy) on emotional prosody.

Keywords: emotional prosody, vocal expression of emotions, speech disorders, hearing loss, dysarthria, infantile cerebral palsy.

INTRODUCTION

Emotions are an inalienable attribute of human behavior: their manifestations can be observed inter alia in speech prosody. A biologically normal person acquires knowledge and skills at the suprasegmental level of language in a natural way during interactions. There are, however, persons affected with disorders in consequence of which learning of prosodic structures of the language or the utilization of acquired knowledge in this area is difficult or impossible. We might suppose that, consequently, the feelings of these persons will manifest differently. The exact and objective description of emotional prosody enables verification of this hypothesis, which I sought to accomplish in this study.

* I would like to extend my gratitude to Prof. S. Grabias, Dr A. Lorenc, and Dr M. Wysocka for methodological support during my research.

THE STATE OF KNOWLEDGE ON THE SUBJECT

Prosody refers to the suprasegmental structure of utterances. Two kinds of prosody are distinguished: linguistic, which complements the linguistic message, and affective, which enables insight into the sender's inner experiences (Gurański et al. 2008).

The subject of the present article is emotional prosody – a phenomenon still insufficiently described and difficult to investigate. While language is a predictable structure that can be examined in intersubjective terms, the experiencing and showing of feelings, despite common biological and social patterns, remain in the sphere of individual behavior. Technological development provides access to objective and more accurate research tools. Despite the fact that phonetics has made huge progress in the field of studies on the suprasegmental level of the language, it took place mainly in descriptions of linguistic prosody (cf. Trochymiuk 2008; Wysocka 2012).

Most studies on emotional prosody produced so far aimed to make a qualitative comparison between acoustic features of emotionally unmarked utterances and affectively marked ones (cf. Murray, Arnott 1993; Sobin, Alpert 1999). Few publications take speech disorders into consideration, being rather concerned with selective phenomena such as investigations of the speech of patients after neurological incidents (cf. K. Gurański et al. 2008, cf. Grabowska 2002; 65–66), or with studying persons suffering from mental diseases (cf. Hoekert et al. 2007, 2002). The foregoing publications are multiple case studies without reference to a general norm because none has ever been defined.

In the field of affective prosody there is also a group of studies conducted for the needs of speech technology. Despite the application of objective research tools, the previous conclusions arising from the works of scholars examining this phenomenon are still relative rather than absolute. The investigations cover only normative utterances. Their results are practically applied in creating programs for automatic recognition of emotions in speech (cf. Igras, Wszolek 2012), or for modeling artificial speech (cf. Demenko 1999).

The abovementioned investigations do not, regretfully, result in proposals for new modern technologies improving the realization of the suprasegmental layer of the language that would be useful in speech therapy treatment. There is no standardized procedure for studying emotional prosody; consequently, it is difficult to compare the results obtained by different scholars. However, the emerging regularities upon which my investigations are based should not be ignored.

THE AIM OF INVESTIGATIONS AND ADOPTED ASSUMPTIONS

The basic research problem in the present study is the question: how do selected emotions manifest in spoken utterances? This outline of the subject suggests specific questions:

1) are there universal acoustic patterns of the vocal expression of particular emotions?

2) what is the impact of diverse speech disorders on the realization of affective prosody?

The goal of my investigations is to characterize and compare manifestations of emotions in spoken utterances of a biologically normal person and of persons suffering from speech disorders of different origin. The conclusions allow me to estimate the impact of decreased perceptual efficiency deficits and articulation efficiency deficits on the prosodic contour of an utterance, which will, in future, enable planning appropriate speech therapy treatment focused on improving specific aspects of the speech signal.

It was Darwin who already came to the conclusion that “if the structure of our organs of respiration and circulation had differed in only a slight degree from the state in which they now exist, most of our expressions would have been wonderfully different” (Darwin 1873, 319). I assume therefore that the human anatomical structure and physiology are the biological foundations of speech. Despite its complex anatomical structure, the vocal apparatus can be compared, to put it simply, to a wind instrument, while speech can be treated as an acoustic wave, disregarding its semantic aspect (Tarasiewicz 2003). C. Basztura points out that there are clear analogies between the structure and mechanisms of the sound source, and acoustic parameters of the signal that it has generated (Basztura 1988). By applying these observations to voice analysis we can conclude that anatomical or physiological abnormalities within the speech organs will result in the occurrence of characteristic changes in the acoustic picture.

I assume that by using the tools of acoustic analysis it is possible to objectively and accurately identify and then compare characteristic physical parameters of the speech signal that distinguish emotionally marked sentences from neutral ones. I also assume that under socially and biologically normative conditions of functioning a person learns certain patterns of emotional expression in a natural way by participating in social interactions. The decreased efficiency of speech organs caused by cerebral palsy makes it difficult to use prosody effectively while hearing loss limits both perception (difficulties in receiving the prosodic pattern) and realization (decreased auditory self-control) (Grabias 2007).

On the basis of these assumptions my hypotheses are that in biologically normal persons there are certain common features of vocal expression that differentiate utterances with emotional coloring while affective prosody in persons with speech disorders is different from the norm and shows other properties that are also differentiated for types of pathology. I will seek to verify these hypotheses in the present study.

PROFILES OF THE SUBJECTS

My investigations were a pilot study, which is why three persons took part in it. They were women aged about 23 years. The first is biologically normal, the second with a hearing loss, and the third suffers from infantile cerebral palsy. All the subjects are intellectually normal. The information about them was derived from interviews, observations, and analysis of the available medical and psychological records.

The biologically normal person is a speech therapist-audiologist by education: therefore, she represents highly normative language. Consequently, she also performs the function of the reader in recordings that are the introductory part of the tasks proper to the other subjects.

The second subject has severe congenital prelingual hearing loss (80-90dB). Her main mode of communication is speech assisted by looking at the interlocutor's mouths. She can also use sign language but utilizes this skill only in interactions with some deaf persons.

The third subject suffers from infantile cerebral palsy from birth. The prevailing cerebral palsy symptoms are: paralysis of the lower limbs and muscle spasticity. Athetoid hand movements are also observable, however. There is also a dysarthric speech disorder characterized first of all by difficulties in respiratory-phonatory-articulatory coordination, by tired, hoarse phonation, and inaccurate articulation arising mainly from the excessive muscle tension of the speech apparatus.

RESEARCH TOOLS AND PROCEDURES

The subjects tried to realize semantically neutral utterances in a neutral way (without being emotionally marked) and in an expressive way – differently emotionally marked: with joy, sadness, and anger. I chose the three emotions because of distinctiveness of their expression (both vocal and mimic), clear connotations¹, and contrast between them (in expression and meaning). The realization of the task proper was preceded by a two-phase preparatory stage. The research tool and procedures were developed by myself for the experiment.

The investigation proceeded according to the following stages.

1. General information about what the investigation will consist in.
2. Interview of the subjects on their biological capabilities and social background.
3. Defining of the selected emotions by :

¹ This is based on the conclusions in my BA thesis on the interpretation of names of emotions by persons with hearing loss. The cognitive definitions of joy, sadness, and anger given by the hearing and hearing-impaired persons were more similar unlike those of embarrassment and admiration, which hearing-impaired persons found difficult to characterize, less unequivocal, and sometimes incomprehensible.

- a) discussion of situations in which people feel joy, sadness, and anger.
- b) description of the appearance and behavior of persons in different emotional states.

Objective: exclude different understanding of affective terms, and train the subjects to put themselves in a right mood.

4. Demonstration of the essence of expressive utterances:

- a) reading transcriptions of the first part of recordings, speculation on the situational contexts of the utterances,
- b) listening to the reader saying single sentences in which the prosody and semantics of the utterances carry the same emotionally marked meaning,
- c) repeating of the phrases heard, trying to copy intonation.

Objectives: remind or inform the subjects that the information about the interlocutor's emotions is conveyed both through semantics and prosody. Repeat the expression patterns in specific emotional states.

5. Demonstration of how to execute the task proper:

- a) detailed explanation what the subject's task will consist in,
- b) listening to the reader speaking an example sentence that does not connote an emotional attitude, each time giving the utterance a different emotional coloring through prosody,
- c) attempts to guess which emotion the presented utterance is associated with (not assessed).

Objective: exclude failure to understand the instruction by the subjects.

6. realization of the task:

- a) reading of two sentences with emotionally neutral content,
- b) placing of the phrases in a situational context,
- c) saying the emotionally unmarked sentences, then saying them with joy, sadness, and anger,
- d) recording of the utterances of the subjects,
- e) repeating of recordings of the utterances that the subjects regarded subjectively as realized in an unsatisfying way.

Remarks:

1. All the sentences spoken by the sentence reader were also presented in a parallel written form.

2. In the conversation I used the right facial expressions and gestures as well as pictograms illustrating particular emotions so that the emotion state spoken about was easy to imagine. .

3. The sentences selected for the task were phonetically so constructed as to consist exclusively of open syllables and as great a number of voiced sounds as possible to enable measuring the values of fundamental frequency at each phonetic segment.

4. Utterances were recorded with a directional microphone in the favorable acoustic environment (although, regrettably, not in soundproofed rooms).

RESULTS OF THE ACOUSTIC ANALYSIS OF UTTERANCES AND THEIR INTERPRETATION

The acoustic analysis of the recordings was carried out using the Praat program. For each answer the following graphs were obtained: oscillogram, spectrogram, and the relationships between fundamental frequency and between voice intensity and time. Shown below are the graphs for a sample answer.

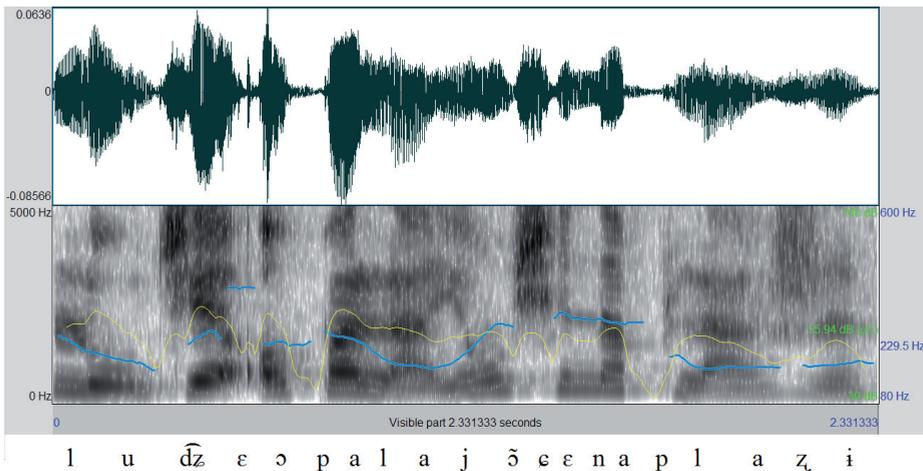


Fig. 1. Oscillogram, spectrogram, intonogram and the chart of intensity of the utterance "ludzie opalają się na plaży" [People are sunbathing on the beach] spoken by the biologically normal person, without being emotionally marked

Using the charts, the following parameters were measured:

1) rate of speech calculated as the quotient of the number of syllables in an utterance and its duration (in seconds). The result is given in the conventional unit syl/s (syllable per second);

2) mean value of fundamental frequency of the whole utterance, measured in hertz (Hz);

3) the range of values of fundamental frequency calculated as the difference between the highest and lowest value F_0 in the utterance. Results given in Hz;

4) the range of speech signal intensity calculated as the difference between the highest and lowest intensity value in the utterance, measured in decibels (dB).

Additionally, the contour of fundamental frequency in all utterances was analyzed in order to describe speech melody, paying special attention to realization of intonemes. I regard an intoneme or the contour of fundamental frequency within

the main stressed syllable (or postictic syllable in justified cases) as the principal functional intonation unit (Wierzchowska, 1980).

The next step was to calculate the arithmetic mean of values of all measured parameters from two sentences for each subject within one emotion.

The abovementioned sets of numerical values were analyzed on two levels with two measures²: expression measure and disorder measure.

The expression measure is intra-individual. It compares the parameter values of emotionally marked sentences in relation to neutral sentences spoken by the same speaker. The measure is calculated using the formula:

$$ME = \frac{E - B}{K} \cdot 100\%$$

in which: ME – expression measure, E – parameter value measured in an emotional utterance, B – parameter value in an emotionally unmarked utterance.

The expression measure makes it possible to objectively compare qualitatively different parameters. It also enables plotting of expression profiles. These are diagrams plotted for each person, specifying values of expression measures calculated for each parameter. It can be determined on their basis how particular emotions manifest in the utterances of individual persons. They also enable us to find whether emotions in persons with speech disorders manifest according to an individual pattern or to the pattern like in healthy persons (even if with certain performance limitations). Presented below are the expression profiles of joy, anger, and sadness plotted for each of the persons investigated.

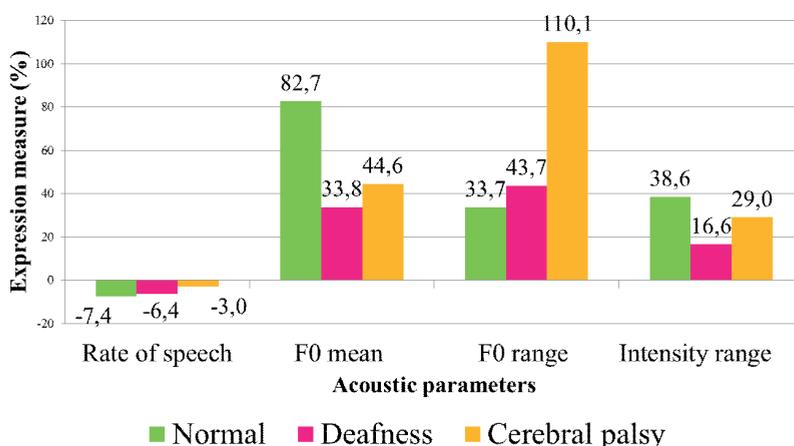


Fig. 2. Joy expression profiles

² The foregoing measures were devised by myself for this study. They are based on elementary mathematical operations.

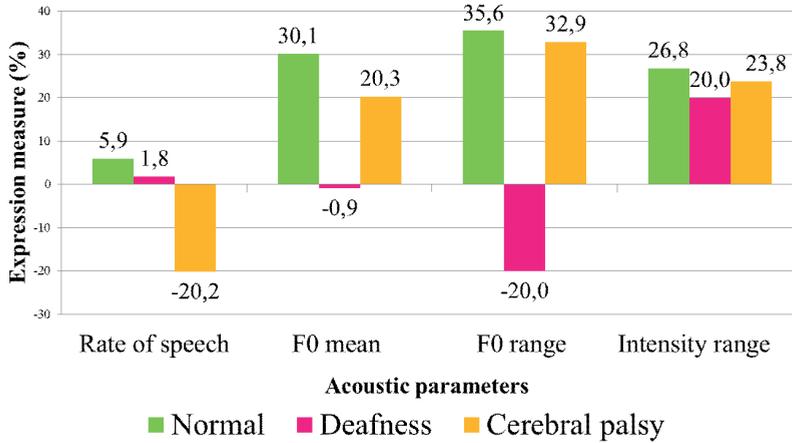


Fig. 3. Anger expression profiles

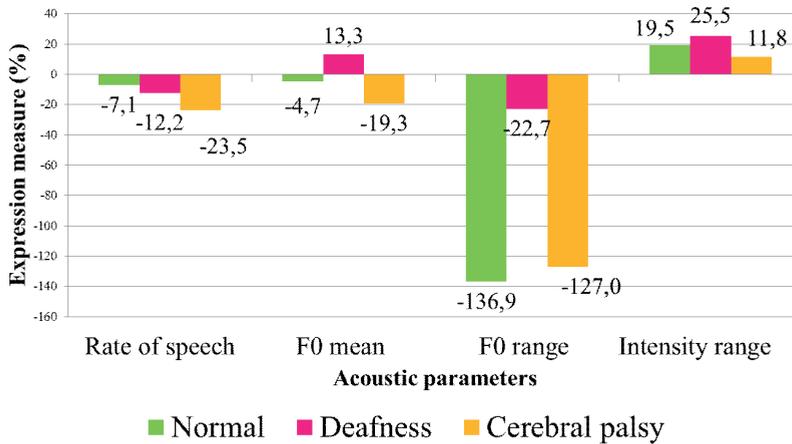


Fig. 4. Sadness expression profiles

The profiles of emotion expression plotted for the persons with speech disorders differ from the normative profiles. In order to assess to what extent these deviations are significant, the measure of disorder should be determined.

The measure of disorder concerns the value of an acoustic parameter within one type of emotion. It is inter-individual: it relates values measured for a person with speech disorders to those obtained for the utterances of a biologically normal person. The measure is calculated using the formula:

$$MZ = \frac{oZ - oN}{oN} \cdot 100\%$$

in which: MZ – disorder measure, oZ – parameter values in the utterances of a person with speech disorders, oN – parameter values in the utterances of a biologically normal person.

The expression measure makes it possible to approximately determine the significance of the identified differences in the way of expressing emotions and to plot disorder profiles or diagrams showing disorder measures calculated for each parameter. Presented below are the disorder profiles for neutral, joyful and sad utterances, and for an utterance spoken with anger:

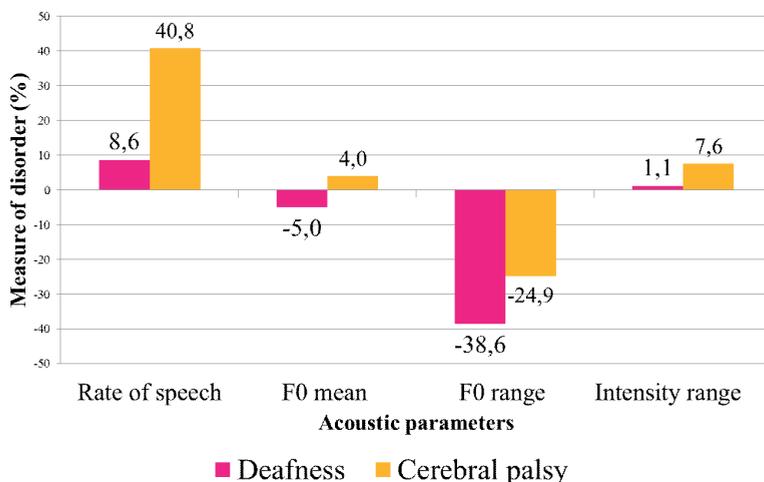


Fig. 5. Disorder profiles: emotionally unmarked

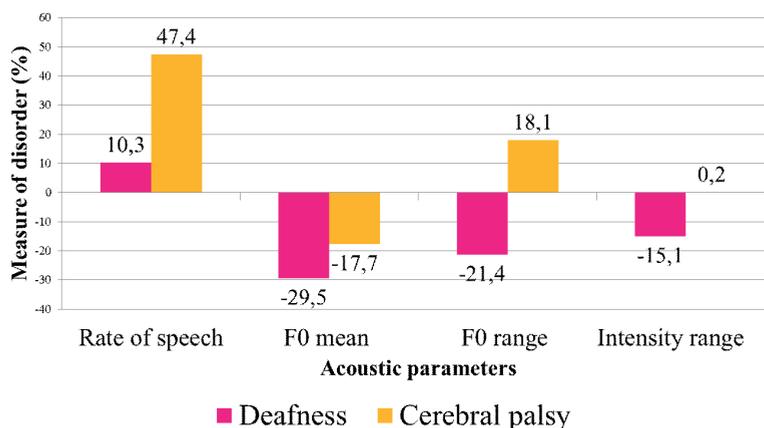


Fig. 6. Disorder profiles: joy

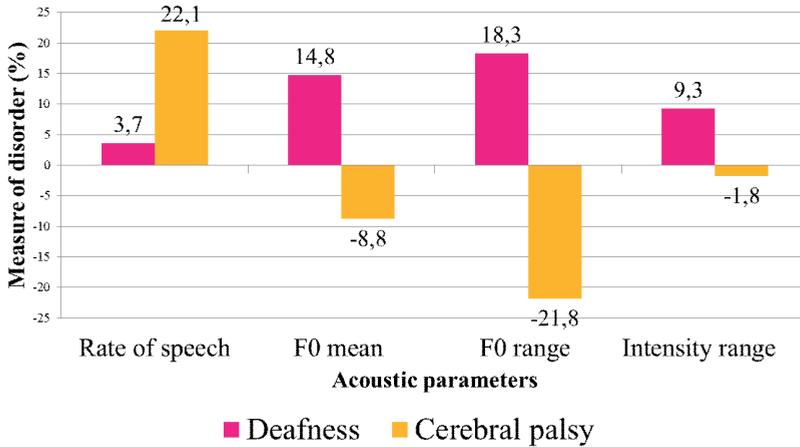


Fig. 7. Disorder profiles: sadness

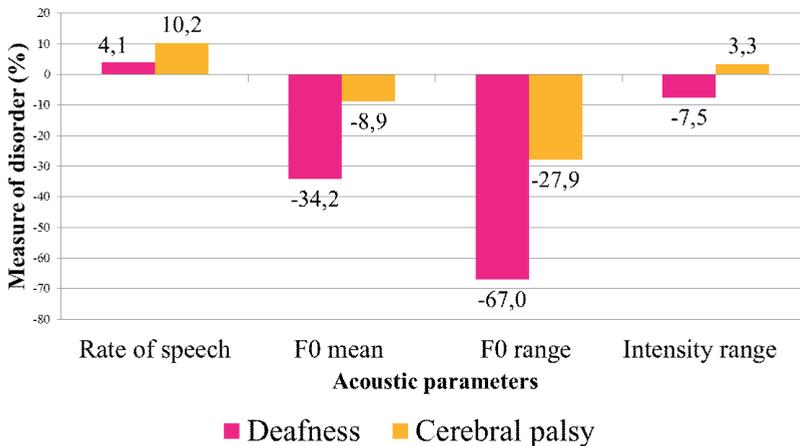


Fig. 8. Disorder profiles: anger

The results of the analysis were thoroughly interpreted. On the basis of the conclusions I tried to construct the acoustic model of manifestations of emotions and to forecast the possibilities of therapeutic management in improving the realization of prosody in persons with speech disorders caused by underdeveloped competence or by limited ability.

INTERPRETATION OF RESULTS, AND CONCLUSIONS

The results of my investigations prove that the acoustic patterns of the vocal expression of particular emotions are characterized by certain regularity. Regard-

less of the kind of emotion, affective utterances are accompanied by increased vocal dynamics. In the expression of sadness the remaining acoustic parameters are lower, while in anger they are higher. Manifestations of joy and anger are similar, yet detailed analysis of their acoustic determinants shows fundamental differences. In a joyful utterance, unlike the one marked with anger, the increase in the values of the acoustic parameters measured is uneven. Moreover, a slowed speech rate is observable whereas its acceleration is a manifestation of anger.

For the expression of each emotion investigated, a dominant parameter can be distinguished in creating an affective impression. In utterances marked by sadness this is a considerably narrowed range of fundamental frequency, in joy – a definitely higher mean value of the laryngeal tone, and in anger – a somewhat accelerated speech rate. These conclusions generally confirm those presented by foreign scholars, and, additionally, specify them in greater detail.

The result of comparative studies on the vocal expression of emotions in biologically normal persons and in persons with speech disorders shows that both a decrease in perceptual efficiency and in articulation efficiency affects the form of emotional utterances. The general conclusion is that because of underdeveloped competence in persons with hearing loss, emotionally marked utterances differ considerably from standard ones, while the limitations caused by dysarthria in infantile cerebral palsy only slightly disturb the production of emotional prosody.

The analysis of utterances of the person with infantile cerebral palsy allows us to say that she knows the expression patterns of particular emotions, and, by overcoming difficulties related to speech disorders, she seeks to produce them using compensation strategies. In expressing joy, the subject finds it difficult to maintain higher fundamental frequency for a long time; that is why she manipulates its range, which yields a similar perceptual effect as a result. However, these frequency fluctuations are always smaller than in the biologically normal person, which disturbs the expression of anger but heightens the expression of sadness.

The affective prosody in the person with hearing loss differs from the pattern to such an extent that emotions may be confused with one another. In acoustic terms, the expression of joy resembles anger, which in turn can be regarded as sadness, and the latter as a neutral utterance. The hearing-impaired subject speaks in a monotonous way but it can be said that to some limited extent she can noticeably change both the mean value of the laryngeal tone and its range. She does it, however, in an individual way, producing different profiles of emotion expression.

Since I studied three persons only, these conclusions cannot be projected on the whole population. Nevertheless, my study has defined a new direction of research on affective prosody, which may in future result in the development of more effective speech therapy methods aimed at improving the realization of suprasegmental linguistic features. These studies would, however, have to be repeated with the participation of a larger group of subjects. They would also

need to be complemented with auditory analysis. It may well also be the case that emotions manifest not only in characteristic prosody, but they also affect the realization of segmental linguistic features, which should be additionally verified.

CONCLUSION

Emotional prosody is widespread in social interactions; that is why its investigation is just as important as description of the segmental elements of the language system. The significant role in the process of language acquisition and efficient communication with the people around us is confirmed by studies in linguistics, speech therapy, or psychology, but also in speech technology or medicine. However, this area of knowledge still remains insufficiently investigated, although the use of advanced research techniques promises further progress in this field.

I would like to express my gratitude to Prof. Dr hab. Stanisław Grabias, Anita Lorenc, PhD and Marta Wysocka, PhD for their methodological support in my investigations.

BIBLIOGRAPHY

- Baszture C., 1988, *Źródła, sygnały i obrazy akustyczne*, Warszawa.
- Darwin K., 1873, *Wyraz uczuć u człowieka i zwierząt*, Warszawa.
- Demenko G., 1999, *Analiza cech suprasegmentalnych języka polskiego na potrzeby technologii mowy*, Poznań.
- Grabias S., 2007, *Język, poznanie, interakcja*, [w:] *Język. Interakcja. Zaburzenia mowy. Metodologia badań. Mowa. Teoria, praktyka*, red. A. Domagała, T. Woźniak., t. 2, Lublin.
- Grabowska A., 2002, *Lateralizacja emocji w mózgu – dane eksperymentalne i kliniczne*, [w:] *Neuropsychologia emocji. Poglądy, badania, klinika.*, red. A. Herzyk, A. Borkowska, Lublin, s. 60.
- Gurański K., Słotwiński K., Podemski R., 2008, *Prozodia mowy w niedokrwiennym udarze mózgu*, „Udar Mózgu”, 10, 2, s. 96–103.
- Hoekert M., Kahn R. S., Pijnenborg M., Aleman A., 2007, *Impaired recognition and expression of emotional prosody in schizophrenia. Review and meta-analysis*, “Schizophrenia Research”, 96, s. 135–145.
- Igras M., Wszolek W., 2012, *Pomiary parametrów akustycznych mowy emocjonalnej – krok ku modelowaniu wokalnejszej ekspresji emocji*, [w:] *Modelowanie i pomiary w medycynie MPM 2012. Materiały XI sympozjum. Krynica Zdrój, 13–17 maja 2012*, Krynica Zdrój.
- Murray I. R., Arnott J. L., 1993, *Towards the stimulation of emotion in synthetic speech. A review of the literature of human vocal emotion*, “Journal of Acoustic Society of America”, 93, 2, s. 1097–1108.
- Sobin C., Alpert M., 1999, *Emotion in speech. The acoustic attributes of fear, anger, sadness and joy*, „Journal of Psycholinguistic Research”, 28, 4, s. 347–365.
- Tarasiewicz B., 2003, *Mówię i śpiewam świadomie. Podręcznik do emisji głosu*, Kraków.
- Trochymiuk A., 2008, *Wymowa dzieci niesłyszących. Analiza audytywna i akustyczna*, Lublin.
- Wierzchowska B., 1980, *Fonetyka i fonologia języka polskiego*, Wrocław.
- Wysocka M., 2012, *Prozodia mowy w percepcji dzieci*, Lublin.