computing@tanet.edu.te.ua
www.tanet.edu.te.ua/computing

ISSN 1727-6209
International Scientific
Journal of Computing

Acoustic Invariant Approach to Speech Sound Analysis for Brand New Speech Recognition Systems (Ukrainian and English)

Maksym O. Vakulenko

Kyjiv Taras Shevchenko National University,
64 Volodymyrsjka, 01033 Kyjiv, Ukraine,
maxvakul@yahoo.com

Abstract: *On the basis of acoustic invariant speech analysis (AISA), the permanent spectral characteristics of the Ukrainian vowels are obtained for various ways of pronunciation including ordinary speech, whisper and changing tone. It is shown that the lowest phonemic frequencies due to vocal fold oscillations or to Helmholtz resonance are not associated with persistent sound features. It is conjectured that the only phonemic invariant is the ratio between formant frequencies, not their absolute values. This analysis is complemented by the computer sound synthesis. We show also that the acoustic invariants of the Ukrainian sound [i] are close to that of English [I]. The results obtained may be useful for specialists in the field of experimental phonetics and speech modelling.*

Keywords: *speech recognition, acoustic phonetics, acoustic invariants, speaker-independent characteristics.*

1. INTRODUCTION

Investigation of human speech sounds is a key task in experimental phonetics and related branches, as well as in various fields of applied science and engineering. Since speech sound analysis is essentially a problem of physics, it can be the scientific basis for linguistic studies in phonetics (see [1-21]), together with associated rules for spelling of borrowed words ([21-24]), foreign language teaching ([25]) and with transliteration systems ([26, 27]) related to questions of International standards.

Satisfactory comprehension of general features responsible for phoneme formation affords one the possibility to construct a contrivance capable of producing and recognizing any speech flow independent of speaker age, sex or emotional condition, as it were a “computational secretary.” This is related to the problem of artificial intelligence. On the other hand, proper understanding of psychological effects on the speech spectra parameters would make it possible to create reliable lie detectors, disease diagnosers and speech-activated locks with selective reaction to certain emotional state. Perhaps, in the future we will be able to simulate at will precise articulation of any given human being, and reproduce the voices of famous singers of the past. In addition, the possibility for automatic audio supplementation of the modern computer dictionaries holds real promise, too.

There are many automatic speech recognition systems presently on the market, which can satisfy more or less whimsical requirements. However, the problems of noise, learning, gender, emotional state, etc. are not yet solved completely, as has been admitted by the well-known experts Wayne Lea [28] and Taras Vincjuk [29]. The most advanced speech software (such as the *Kurzweil* or *Dragon dictate* computer programs) strongly depends on these factors, requires tedious teaching and makes numerous errors – in addition to requiring vast amounts of computer memory for the word storage. By comparison: a dog can easily distinguish its name in a noisy street whoever utters it – either aloud or in a whisper, – and perceives intonation, and is quite quick in learning. On the other hand, modern computers bear such great intellectual potential that the electronic machine "Deep Blue" was able to notch a victory over the chess world champion. So, despite great efforts aimed at constructing sufficiently "clever" automatic recognition speech devices, one must admit that satisfactory success has not yet been achieved – even to the degree of "canine intellect". The language engineers use to involve an excessive quantity of speech parameters (acoustic and articulation ones) thus mixing the invariants with the

variants in sound properties. Thus the mentioned problem still requires further investigation. A truly consistent physical picture of speech sound formation and perception with an account of reliable invariant spectral characteristics has yet to be put forward.

It is remarkable that human speech studies heretofore [1-17] have been rested mostly upon the basis of articulation. Such articulatory emphasis is also inherent in the fundamentally important monograph "Acoustic phonetics" by K. Stevens [1] where the acoustic characteristics of any sound are inferred from the vocal tract configuration. This bias is explained by the fact that the articulation features are more transparent and easier to follow and interpret, and such analysis is less machine-dependent. Being a much more explored field than its acoustic counterpart, the articulatory approach is widely used in speech recognition systems that are based, in most cases, on the so-called generative model [29]. Within this model, any incoming sound signal is compared to a set of machine-generated reference waveforms produced on the basis of knowledge about human articulation. Not surprisingly, this approach makes it impossible to separate speaker-independent acoustic characteristics responsible for formation of a given speech sound from ones reflecting the talker's individuality. In addition, the mechanism of speech sound production, as reported in [14], is too complicated and difficult to be fully specified within existing physical models.

A somewhat different emphasis is observed in human communication where it is acoustic analysis of incoming sound waveforms that is of utmost importance in speech perception. To perceive and understand speech, we have no need to know in detail how every sound is produced. If we were to articulate every speech sound heard in order to understand it, we would be simply unable to perceive more than one speaker at the same time. Actually, we can easily recognize and understand simultaneous speech. Children learn to speak from hearing, where knowledge of articulation serves as an auxiliary tool. It was proved both theoretically and experimentally that speakers tend to produce formant distances that conform to the acoustic targets (see [15]). It is then not surprising why children who are born deaf are also mute: they lack data to carry out the acoustic speech analysis in their brains. This is why the speech of deaf persons trained on an articulation basis, can never be perfect. And this is also how parrots learn to speak: their articulatory apparatus is quite different from that of *Homo sapiens*, so that they must take advantage of acoustic analysis. This sort of analysis underlies the general speech recognition mechanism that occurs everywhere in Nature. So the acoustic side of the problem which relies on wave analysis of uttered speech, is of prior importance for further progress here.

In this paper, we will be concerned with acoustic characteristics of incoming speech waveforms, in regard to their major importance for speech perception and recognition. As we noted earlier, any human speech sound contains both acoustic characteristics of the speech sound itself, as well as information about the special features of the talker. Acoustic invariant speech analysis (AISA) methodology is developed here to make it possible to separate these.

We will elucidate the physical picture of the speech sound formation and analyze the spectral characteristics of Ukrainian vowels for different manners of pronunciation including ordinary speech, whisper and changing tone (singing). The formants that appear in these sounds will be considered to fall into two separate classes: the class of principal (essential, chief, cardinal) formants that correspond to invariant acoustic characteristics, and class of collateral (incidental, occasional) formants associated with special features of the individual or a given utterance (such as timbre, sound interaction, etc.). It will be shown that in normal speech, the lowest phonemic frequencies due to vocal cord oscillations or Helmholtz resonance, cannot be treated as the essential formants (though in high-tone speech, the main frequency may reach the level of the first formant: $F_0 = F1$). Additional resonances may arise for high enough tone frequencies when a resonator can accommodate a number of higher overtones: $n = 1, 2, \ldots$, etc. In this case, there appear additional formants with the frequencies proportional to those of the principal ones. Most importantly, it is conjectured that the only phonemic invariant is the ratio of cardinal formant frequencies. This analysis is complemented by computer sound synthesis. We show that the acoustic invariants of the Ukrainian sound [i] are close to those of English [**I**]. If one sound is "mixed" with another, additional prominences arise that have a ratio corresponding to a mixed sound. The results obtained may be useful for specialists in the field of experimental phonetics and speech modelling. The acoustic invariants found for the vowels analyzed, should be taken into account in modern talker-independent speech recognition software.

2. Acoustic invariant speech analysis: background and methodology

To begin with, let us recall that in the human speech tract, two kinds of resonance are possible: a Helmholtz (low-frequency) resonance that appears

in a volume with a narrow constriction, and a tube resonance, where the tube may be closed or half-closed (see [1]). For a configuration consisting of a large volume (that gives rise to an acoustic compliance) and a narrow constriction (acoustic mass) where the dimensions of the volume and the constriction are small compared to the wavelength $\lambda_H = c/f_H$, the Helmholtz natural frequency is

$$f_H = [c/(2\pi)][A/(Vl)]^{1/2} , \quad (1)$$

where A and l are the cross-sectional area and length of the narrow tube, respectively, and V is the volume of the large tube (for more details, see [1: 142]). This resonance may exist at low enough frequencies (up to about 500 Hz), where higher harmonics are not involved.

The natural frequencies, or eigenfrequencies, for a half-closed tube of length l and uniform cross-sectional area are:

$$f_n = [(2n - 1)/4](c/l), \quad (2)$$

where $n = 1, 2, 3, ...$ is the natural frequency number (see details in [1: 138-139]).

If a tube is closed at both ends,

$$f_n = [(n - 1)/2](c/l). \quad (3)$$

In high-tone speech when the frequencies are high enough to eliminate Helmholtz resonance, only (2) and (3) are valid, where the parameter n runs higher values.

Let us make a necessary distinction between the natural frequencies of the vocal tract – they are fixed for a given tract configuration of a given person – and the formant frequencies that determine the given speech sound through corresponding acoustic modes (this difference is not emphasized in [1]). The former quantities are, of course, the speaker-dependent ones. If the frequency of any partial harmonic in the source spectrum (including the fundamental frequency itself) is close to some of the vocal tract eigenfrequencies, its amplitude grows. This results in the formation of the desired (or haphazard) speech sound. Should we manage to elucidate the way in which the prominences in any individual speech sound is connected with the talker-independent factors, this will provide significant help in solving the problem of speech recognition.

Let us emphasize that we avoid such senseless definitions of formants as "zones of frequency increase" [16: 42]: increase in frequency ought to be distinguished from increase in amplitude. In addition, we will see later that a harmonic that is crucial for a given sound, is not necessarily larger than the others (this happens when a Helmholtz resonance occurs, and when one sound is "mixed" with another, and if any sound is influenced by its neighbour). That is why we regard a formant as an acoustic mode that contributes to the speech sound given, and why we will need to keep this statement in mind in what follows.

Let us examine now special features of the Helmholtz resonance displayed in the incoming speech sound wave, in the context whether or not it may be relevant to persistent speech sound characteristics.

1. The air pressure oscillations coming into the human ear or a microphone membrane, are described by the well-known expression

$$p_r(f) = [if\rho/(2r)]U_r(f)exp(-2\pi ifr/c), \quad (4)$$

where i is an imaginary unity, f is the frequency, ρ is the air density, c is the velocity of sound in the air, r is the distance from the mouth, p_r is the sound pressure produced at distance r, U_r is the volume velocity at distance r (see, for example, [1: 127-128]). It is seen from (4) that the low-frequency stimulus results in smaller pressure amplitude than the high-frequency one. That is, the higher frequencies are more distinguished in the transmitted waveform. This indicates that the most important acoustic characteristics of speech sounds are likely to reside in the high-frequency range.

2. The Helmholtz resonance in the vocal tract is usually manifested by broad bandwidths associated with reduced amplitudes, due to large energy dissipation in the low-frequency range. This energy loss is conditioned by: impedance of the vocal tract walls ([1:157,193; 7, 8]), viscosity and heat losses ([2, 3], see also [1: 160-161]), glottal opening ([1: 165-166]) or oral cavity constriction ([1: 534]), and nonlinear acoustic resistance (1: 163-164]). This reduces the detectability of corresponding amplitude and henceforth minimizes its role in sound formation. For most vocal tract configurations in whisper speech, the acoustic losses due to wide glottal opening are so large that the critical damping of Helmholtz resonance occurs (see [1: 165,171] and [5, 6]). In these cases, the relevant spectral prominence is completely dissipated. Nevertheless, such sounds as [i], [u], [l], [m], [n], [s] whose first formant frequencies are conventionally believed to be caused by Helmholtz resonance, do not vanish: they are still heard and recognized even when whispered.

3. The fleshy surfaces of the tongue, cheeks, and pharynx are not rigid, so the mass reactance of the walls gives rise to significant change in the lowest natural frequency of the vocal tract. As measured in [7, 8], this correction is about 180 Hz for adult male talker and about 190 Hz for female. As a result, such

frequencies are the most difficult to manipulate by articulatory movements. It was stated in [1: 160] that the frequency of 250 Hz is only one-half as sensitive to changes of the vocal tract configuration. This indicates again that Helmholtz frequencies play virtually no major role in the formation of specific features of a given sound.

4. The vowel perception experiments [11, 12] show that a cluster of lowest harmonics is interpreted as a single prominence if they are sufficiently close to the fundamental frequency – approximately not farther than 300 Hz in the range below 500 Hz.

However, a human ear can grasp much subtler frequency differences. Musicians and singers, for example, can distinguish quartertone intervals that correspond to several Hertz or less in the speech frequency range. Hence, albeit duly heard, some of the lowest frequencies are merely omitted by our brain when it analyses such wave. As far as the tone pitch that is involved in such important phenomena as intonation and singing, is determined by the fundamental frequency, the latter cannot be disregarded. Consequently, it is the Helmholtz frequencies that are likely treated by our brain as the least important for sound discrimination.

5. Let us recall that the fundamental frequency F_0 determining the tone pitch is the lowest one in the sound wave. No prominence can arise below it. Besides, this frequency can easily reach 500 Hz in normal discourse, and 1000 Hz or higher during singing while the Helmholtz range lies below about 500 Hz. Nevertheless, in the course of such high-tone speech or singing, the sounds [i], [u], [l], [m], [n], [s] that are classically believed to have their first formants in the range 200-400 Hz, do not fade away. This leads to the conclusion that the low-frequency behaviour has only minor effects on the sound quality.

To summarize this criticism, we should acknowledge that the Helmholtz effects are either vague or completely absent in most pronunciation modes, and that this resonance could not give rise to regular and distinct features of speech sounds. Consequently, it may only contribute – for low enough fundamental frequencies – to collateral characteristics describing the speaker's individuality such as timbre or to those things associated with sound interaction.

To separate the invariant acoustic characteristics, let us address the experimental fact that playback of recorded speech at enhanced speed ("Buratino voice") does not result in the sound transmutation. In other words, multiplication of each formant frequency by the same factor preserves the distinctive features of the given sound: [s] remains [s] (it is not transformed into [f], for example), [i] remains [i] (not [u] or [l]), and so on. Such playback is equivalent to excitation of higher harmonics of the fundamental wave producing resonance in a tube (we remember that the Helmholtz resonance is not possible for overtones with higher numbers).

Thus one may expect that the invariant characteristics of the speech sounds are relative – not absolute – quantities, and that they are induced by the tube resonance. This relativity conjecture is indirectly supported also by results of [17] where some correlation between F_0 and vowel formant frequencies was reported, and by those of [18] where microdynamic behaviour of vowels in French, English and Czech was not found to depend on fundamental frequency. Such invariants are determined by the ratio between the main formant frequencies. By using a ratio of formant frequencies we can account for the important fact that male and female talkers with differing lengths of their vocal tracts are nevertheless perceived to be producing the same sound, even though their absolute values for F1 and F2 might differ some.

This conjecture is a rational explanation of the puzzle put forward by R. Feynman [30]: why some musical intervals are perceived as pleasant ones? They just reproduce the main intrinsic features of human speech. To make a ratio, two parameters are needed. In view of this, each non-nasal speech sound is expected to possess two independent principal formant frequencies (usually corresponding to back and front cavity resonances), and the nasals have three such frequencies (the additional one comes from the nasal cavity). This assumption correlates well with the experimental data [13] showing that only two formants are required for satisfactory perception of non-nasal vowels.

The experimental part of this work is aimed not only to collect some acoustic data but, most importantly, to argue for our theoretical prediction that the invariant speech sound parameters are the relative ones, and that they cannot arise from the Helmholtz resonance. To carry out acoustic invariant speech analysis (AISA), we will seek such stable formants across different speakers, various ways of pronunciation, and alternative wave analysis procedures, and calculate corresponding ratios.

3. Speech sound analysis and synthesis

Four native Ukrainian speakers were involved in our experiments: two women (19 and 26 years old) and two men (both 34 years old). They pronounced (in triple repetition) separate sounds in normal tone of voice, in whisper and in changing tone, and words and given word combinations ("khata", "kha-kha", "ghadaty", "khodyty", "sydity", "lezhaty", "ja pidu ghuljaty", "byky", "boky", "jakby", "vyjty", "uvijty",

"za Sybirom sonce skhodytj", "zillja", "shyttja", "zhinochi khytroshhi", "dokir", "zdijmaty", "klepka", "leghko", "khtosj", "ne treba", "ce bude", "jak z krolykamy") in normal tone and whisper. This procedure allows one to filter out the characteristic independent on the way of pronunciation. We are not concerned in accumulating vast amount of acoustic data from a large number of talkers. The key point here is taking advantage of different pronunciation modes, where the normal speech is the 1st one, whisper is 2nd, rising and falling tone are the 3rd and 4th, respectively. Should the same acoustic parameter be present within all these modes, it may be regarded as a possible speech sound invariant. In this sense the acoustic data within these pronunciation modes obtained from one speaker, are far more representative than in the combination "normal speech of four subjects".

The speech signals were recorded in an isolated room with a unidirectional dynamic microphone Tech TDM-204 at a 5-10 cm distance from the subject's lips using the computer program Sound Forge 4.0, and analyzed with the programs WaveLab 2.1 and CoolEdit 95. The fundamental and formant frequencies were measured manually on the oscillogram and 3D frequency analysis spectrogram obtained by WaveLab 2.1 and 2D spectrogram by CoolEdit 95, then compared and averaged. Statistical errors are determined by the human factors, wave analysis accuracy inherent to the programs, precision step of the spectrogram, and the bandwidth tail.

We were concerned here with the Ukrainian vowels [a], [o], [u], [i], [y], [e] (we use the transliteration by the Ukrainian Latinics [26, 27] whenever addressing the Ukrainian alphabet).

Essential wave amplitudes in the region 200-400 Hz for [y] and [u] arising from the Helmholtz resonance, were observed in normal discourse. However, these prominences were absent for high enough fundamental frequency when the first overtone was higher than the Helmholtz frequency: $f_1 = 2F_0 > f_H$, and, in most cases, for whispered speech. This supports our earlier conclusion that the Helmholtz resonance cannot give rise to permanent, invariant features of speech sounds.

When the pitch of [a] is lowering, some tone harmonics fall into the resonance regions – in good agreement to the picture described by R. Feynman [30]. In such case, the next overtone goes into resonance and the previous one is damped. Naturally, these resonance attributes are absent in the main tone of [i]. See [19-21] for relevant spectrograms.

Table 1 presents a summary of invariant frequency characteristics of the Ukrainian vowels: mean values of the formant frequencies together with the relevant resonance zones, and ratios

Table 1. Invariant Acoustic Characteristics Of Ukrajinian Vowels

Sound	Speaker	Mean value of Fp1 and its resonance zone, Hz	Mean value of Fp2 and its resonance zone, Hz	Ratio Fp2/Fp1
[a]	1	830 *(760–900)*	1100 *(1050–1150)*	4/3 (1.33 ± 0.04)
	2	830 *(800–850)*	1100 *(1000–1200)*	4/3 (1.33 ± 0.04)
	3	820 *(750–880)*	1100 *(1000–1200)*	4/3 (1.34 ± 0.04)
	4	820 *(750–900)*	1100 *(1000–1200)*	4/3 (1.33 ± 0.04)
[o]	1	530 *(480–580)*	790 *(710–880)*	3/2 (1.49 ± 0.15)
	2	550 *(510–700)*	820 *(720–1000)*	3/2 (1.49 ± 0.15)
	3	530 *(470–680)*	790 *(700–880)*	3/2 (1.49 ± 0.15)
	4	530 *(470–700)*	790 *(700–1100)*	3/2 (1.49 ± 0.15)
[u]	1	3600 *(3400–4000)*	5800 *(5400–6100)*	5/3 (1.61 ± 0.08)
	2	4100 *(3800–4300)*	7000 *(6400–7600)*	5/3 (1.71 ± 0.06)
	3	3700 *(3400–4100)*	6100 *(5500–6500)*	5/3 (1.65 ± 0.09)
	4	3800 *(3600–4000)*	6400 *(6200–6600)*	5/3 (1.68 ± 0.04)
[y]	1	1800 *(1300–2000)*	2100 *(1800–2400)*	6/5 (1.17 ± 0.07)
	2	1800 *(1600–2100)*	2200 *(1900–2600)*	6/5 (1.22 ± 0.06)
	3	1800 *(1400–2000)*	2200 *(1800–2700)*	6/5 (1.22 ± 0.06)
	4	1800 *(1500–2100)*	2200 *(1900–2500)*	6/5 (1.22 ± 0.06)
[i]	1	2300 *(2000–3000)*	2900 *(2500–3800)*	5/4 (1.26 ± 0.08)
	2	2400 *(2000–3500)*	3000 *(2500–4400)*	5/4 (1.25 ± 0.07)
	3	2500 *(2100–3400)*	3100 *(2700–4200)*	5/4 (1.24 ± 0.08)
	4	2400 *(2100–3400)*	3000 *(2500–4200)*	5/4 (1.25 ± 0.07)
[e]	1	700 *(610–780)*	2100 *(1600–2400)*	3 (3.00 ± 0.30)
	2	700 *(580–950)*	2100 *(1800–2800)*	3 (3.00 ± 0.30)
	3	700 *(550–930)*	2100 *(1700–2500)*	3 (3.00 ± 0.30)
	4	700 *(520–920)*	2100 *(1600–2600)*	3 (3.00 ± 0.30)

between second and first formant frequencies. Speaker 1 was female, aged 26, height 165 cm, talker 2 female, aged 19, height 168 cm, speaker 3 male, aged 34, height 176 cm, talker 4 male, aged 34, height 175 cm.

As expected, the acoustic invariants are conditioned by the tube resonance. The Helmholtz resonance that gives rise to frequencies in the range 300-500 Hz in the sounds [i], [y], [u], is not so distinctive in whispered speech and completely vanishes in high-pitched sounds.

The number of harmonics is minimal in [i] and maximal in [a] and [o]. The first overtone has the largest amplitude in [u] and [y]. The most distinctive resonance (amplitude contrast) is displayed in [a] and [o], whereas the sound [y] is the most "vague" one.

In normal speech, the spectrum of [u] is dominated by the low-frequency Helmholtz resonance where $F1_H$ = 300 – 450 Hz and $F2_H$ = 500 – 780 Hz with r_H = 5/3 (1.70 ± 0.03). In high-tone speech and in whisper when restrictions on this kind of resonance are imposed, there dominate relatively high-frequency prominences of formant pairs with the same invariant ratio, r = 5/3 (presented in Table 1). In some cases we observed also and additional formant with F_{add} = 2100 – 2600 Hz where $Fp1/F_{add}$ = 5/3. During whisper utterance, we observed also an extra pair $Fp1_{ex}$ = 3000 – 32000 Hz and $Fp2_{ex}$ = 5000 – 5200 Hz with r = 5/3. Most likely, all these frequencies are conditioned by tube resonance in one of vocal tract cavities and in constriction between two cavities.

When the tone pitch is increased, the corresponding formant pairs move higher in the frequency scale. For example, we observed the main pair of principal formants *Fp1* = 3000 Hz (2800–3200 Hz), *Fp2* = 4000 Hz (3800–4200 Hz) and the additional one *Fp1*(2) = 6000 Hz (5700–6300 Hz), *Fp2*(2) = 8000 Hz (7700–8300 Hz) corresponding to a high-pitched sound [a] – with the main tone frequency F_0 = 800 Hz (well above the Helmholtz range). The formant ratio remains constant here: r = 4/3.

This constancy of the formant ratio is a very important result confirming our early theoretical prediction that the only phonemic invariant is the relative quantity of cardinal formant frequencies.

If one sound is "mixed" with another (such as [u] in the combination [ou] or in the o-type utterance), additional prominences arise that have a ratio corresponding to the mixed sound. In particular, the sound [u] with the "o-impurity" was observed to acquire additional formants F1(o) = 2400-3000 Hz, F2(o) = 3000-4500 Hz with the ratio r(o) = F2(o)/F1(o) = 3/2.

The analysis performed allows one to create any given sound by means of the wave production. We emphasize that artificial sounds (synthesized waveforms [a], [i], [y]) have much more stable and clear audio performance than natural ones.

The tube resonance formants of various acoustic realization of the English phoneme /i/ ([i:] and [**I**]) were found to have relatively stable frequencies: *Fp1* = 2000 Hz (1800 – 2200 Hz), *Fp2* = 2500 Hz (2300 – 2700 Hz), with a formant ratio *r = Fp2/Fp1* = 5/4 (large tertian) that corresponds to the Ukrainian [i]. The low-frequency (~300 Hz) incidental formant caused by the Helmholtz resonance cannot be involved in any invariant ratio.

Significant variation in the harmonic amplitudes across talkers was observed that correlates well with the experimental data [5, 6, 9, 10] stating that the amplitude of the harmonics near 2.5 kHz relative to the amplitude of the first harmonic can vary as much as 20 dB across speakers. This fact allows one to assume that the voice timbre of a given person is manifested in individual relations between different overtones.

4. Summary

We have analyzed spectral characteristics of Ukrainian vowels for different ways of pronunciation including ordinary speech, whisper and changing tone. On the basis of acoustic invariant speech analysis (AISA), acoustic invariants of these sounds were found. It is assumed that the speech sound formants fall into two classes: principal (essential, chief, cardinal) formants corresponding to invariant acoustic characteristics and collateral (incidental, occasional) formants associated with individual or utterance special features. It has been shown here that in normal speech, the lowest phonemic frequencies due to Helmholtz resonance, give rise to occasional formants. Additional resonances may arise for high enough tone frequencies when a resonator can hold a number of higher overtones: *n = 1, 2, …*, etc. In this case, there appear additional formants with the frequencies proportional to those of the principal ones. We show also that the acoustic invariants of the Ukrainian sound [i] are close to that of English [**I**].

The most important finding of this work states: the only phonemic invariant is the ratio between formant frequencies, not their absolute values. The results obtained may be useful for specialists in the field of experimental phonetics and speech modelling. The obtained acoustic invariants should be taken into account in modern talker-independent speech recognition systems.

5. References

[1] K.N. Stevens, *Acoustic Phonetics*. MIT Press, 1998. 607 p.

[2] G. Fant, *Acoustic theory of speech production*. The Hague, Netherlands, 1960.

[3] J.L. Flanagan, *Speech analysis, synthesis, and perception*. Berlin: Springer-Verlag, 1972.

[4] H.M. Hanson, "Glottal characteristics of female speakers: Acoustic correlates," *J. of the Acoustical Soc. of America, 101*, 1997, pp. 466-481.

[5] K.N. Stevens, and H.M. Hanson, "Classification of glottal vibration from acoustic measurements." In: *Vocal fold physiology*: Voice quantity control / O. Fujimura, M. Hirano, Eds. San Diego: Singular, 1995. Pp. 147-170.

[6] H.M. Hanson, *Glottal characteristics of female speakers*. PhD dissertation, Harvard University, Cambridge MA, 1995.

[7] O. Fujimura, and J. Lindqvist, "Sweep-tone measurements of vocal-tract characteristics," *J. of the Acoustical Soc. of America, 49*, 1971, pp. 541-558.

[8] G.Fant, L. Nord, and P. Branderud, "A note on the vocal tract wall impedance," *Speech Transmission Laboratory Quarterly Progress and Status Report 4*, Royal Institute of Technology, Stockholm, Sweden, 1976. Pp. 13-27.

[9] E.B. Holmberg, R.E. Hillman, and J.S. Perkell, "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice," *J. of the Acoustical Soc. of America, 84*, 1988, pp. 511-529.

[10] D.H. Klatt, and L.C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. of the Acoustical Soc. of America, 87*, 1990, pp. 820-857.

[11] H. Traunmueller, "Perceptual dimensions of openness in vowels," – *J. of the Acoustical Soc. of America, 69*, 1981, pp. 1465-1475.

[12] K.A. Hoemeke, and R.L. Diehl, "Perception of vowel height: The role of F1-F0 distance," – *J. of the Acoustical Soc. of America, 96*, 1994, pp. 661-674.

[13] R. Carlson, B. Granstroem, and G. Fant, "Some studies concerning perception of isolated vowels," *Speech Transmission Laboratory Quarterly Progress and Status Report 2-3*, Royal Institute of Technology, Stockholm, Sweden, 1970. Pp. 19-35.

[14] Zh. Zhang, J. Neubauer, and D.A. Berry, "The influence of subglottal acoustics on laboratory models of phonation," – *J. of the Acoustical Soc. of America, 120*, 2006, pp. 1558-1569.

[15] L. Menard, J.-L. Schwartz, L.-J. Boë, and J. Aubin, "Articulatory-acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model," *J. of Phonetics*, 35, 2007, pp. 1-19.

[16] N.I. Tocka, *Vowel phonemes of Ukrainian literature language*. Kyjiv, Kyjiv University publishing house, 1973, 193 p. (in Ukrainian).

[17] T.M. Nearey, and P.F. Assmann, "Information conveyed by f0 for vowel identification," *J.* of *the Acoustical Soc. of America, 119*, 2006, pp. 3339.

[18] T. Dubeda, and E. Keller, "Microprosodic aspects of vowel dynamics – an acoustic study of French, English and Czech," *J. of Phonetics, 33*, 2005, pp. 447-464.

[19] M.O. Vakulenko, "Analysis and synthesis of the sound spectra of human speech," *Pulsar*, № 6-7, 1999, pp. 20-23.

[20] M.O. Vakulenko, "Acoustic characteristics and invariants of the Ukrainian sounds", *Scholarly News of the KSLU UNESCO Chair*. Philology, Pedagogics, Psychology. Vol. 1. Kyjiv, 2000, pp. 62–66 (in Ukrainian).

[21] M.O. Vakulenko, and O.V. Vakulenko, "Ukrainian spelling: view from Ukraine". *Scientific reports of the Higher School Academy of Sciences of Ukraine*. Vol.4. Kyjiv-Khreshhatyk, 2002, pp.129-138 (in Ukrainian).

[22] Maksym Vakulenko, *Russian-Ukrainian Dictionary of Physical Terminology* / Prof. O.V. Vakulenko, Ed. Kyjiv, 1996, 236 p. (in Ukrainian).

[23] Maksym Vakulenko, *On the "difficult" problems of Ukrainian spelling*. Kyjiv, "Kurs", 1997. 32 p. (in Ukrainian).

[24] Maksym Vakulenko, "Spelling aspects of the terminology as a science". *Book Chamber News*. Kyjiv, 1998, #11, pp.15-17.

[25] *Studies in Communicative Phonetics and Foreign Language Teaching Methodology* / M.P. Dvorzhetska, A.A. Kalita, Eds. K., *Lenvit*, 1997.

[26] M.O. Vakulenko, "Transliteration Through a Slavonic Latin Alphabet: Saving Information and Expenses," *Kyjiv Linguistic University News*. – Philology. – V.2, №1, 1999, pp. 85-94.

[27] Maksym Vakulenko, "Ukrainian Latin alphabet as Standardized Addition to Ukrainian Orthography". *Library News*, 1998, #2, pp.10-12 (in Ukrainian).

[28] *Trends in Speech Recognition* / Wayne A. Lea, Ed. – Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1980.

[29] T.K. Vincjuk, *Analysis, perception and interpretation of the speech signals*. Kyjiv, 1987, 262 p. (in Russian).

[30] *The Feynman lectures on physics* // Richard P. Feynman, Robert B. Leighton, Matthew Sands. Vol.1. Addison-Wesley publishing company, 1963.

Dr. Maksym O. Vakulenko *was born in Kyjiv (Kiev), Ukraine, October 17, 1964. He is Associate Professor at the English Phonetics Chair of the Kyjiv National Linguistic University, Head of the Experimental Phonetics Laboratory of the Kyjiv National Linguistic University, and Senior researcher at the Kyjiv Taras Shevchenko National University since 2004.*