Discrete & Continuous Models & Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2025, 33 (3) 309–326
http://journals.rudn.ru/miph

Research article

UDC 004.9+086.7+534+621.397+681.84/.85+791.43+811.58
PACS 01.50.F-, 02.60.-x, 29.85.-c, 42.30.Va, 43.38.Md, 43.58.+z, 43.58.Kr, 43.60.+d, 95.75.Pq
DOI: 10.22363/2658-4670-2025-33-3-309-326 EDN: HEVOIT

Research of hieroglyphic signs using audiovisual digital analysis methods

Maia A. Egorova, Alexander A. Egorov

RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

(received: February 10, 2025; revised: March 1, 2025; accepted: April 20, 2025)

Abstract. A study of ancient written texts and signs showed that the hieroglyphs and structure of the archaic sentence have much in common with the modern Chinese language. In the context of the history and evolution of the Chinese language, its characteristic tonality and melody are emphasized. The main focus of the work is on the study of the sound properties of hieroglyphs (keys / Chinese radicals) found simultaneously in ancient inscriptions as well as in modern text messages. The article uses modern digital methods of sound analysis with their simultaneous visualization. To characterize the sound of hieroglyphs (in accordance with the Pinyin phonetic transcription adopted in China), two (FI, FII), three (FI, FII, FIII) or four (FS, FI, FII, FIII) formants are used, which create a characteristic F-pattern. Our proposed model of four formants for typical hieroglyphs is called the basic one "F-model", it's new and original. To visualize the formants, digital audio signal processing programs were used. The data obtained were compared with the corresponding spectrograms for Mandarin (standard) Chinese. Their correspondence to each other has been established. When analyzing F-patterns, an original model was used, which made it possible to characterize spectrograms in the frequency and time domains. The formalized description of basic components of basic "F-model" of a pronunciation of hieroglyphs is given. In conclusion, several areas are noted in which the use of various methods of audiovisual research is promising: advanced innovative technologies (artificial intelligence and virtual reality); television, theatrical video production; evaluation of the quality of audiovisual content; educational process. The present study has shown that described promising research methods can be useful in analyzing similar ancient hieroglyphs.

Key words and phrases: speech analyzing, Chinese hieroglyphs, spectrogram, formants, subharmonics, basic "F-model", linguistics, data processing, computer modeling

For citation: Egorova, M. A., Egorov, A. A. Research of hieroglyphic signs using audiovisual digital analysis methods. *Discrete and Continuous Models and Applied Computational Science* **33** (3), 309–326. doi: 10.22363/2658-4670-2025-33-3-309-326. edn: HEVOIT (2025).

1. Introduction

Sound waves are an example of an oscillatory process. The simplest sound wave is a periodic (i.e., the amplitude values are repeated at regular intervals) oscillation described by a sinusoid. The sound is characterized by a number of parameters (see e.g. [1, 2]). The human vocal cords produce complex sounds. The presence of complex vibration movements, which form both the main tone and overtones, is one of the reasons for the appearance of complex sounds [2].

© 2025 Egorova, M. A., Egorov, A. A.

This work is licensed under a Creative Commons "Attribution-NonCommercial 4.0 International" license.

The history of the Chinese language has more than three thousand years [3–7]. The Chinese language, the Chinese (Sinitic) family, is a language branch that includes language and dialect groups united by a common script. In its standard form, Chinese is the official language of the PRC and Taiwan, and one of the six official and working languages of the United Nations. Modern Chinese language Putonghua was created artificially in the middle of the 20th century. It is based on the vocabulary and grammar of "Mandarin" (Mandarin Chinese or Northern Chinese) and Beijing dialect is the source for pronunciation, i.e. Putonghua phonetics and vocabulary is based on the pronunciation norm of the Beijing dialect, which belongs to the northern group of dialects of Chinese language [4–9].

The modern Chinese language Putonghua (普通话) is tonal, that is, each syllable that has an accent is pronounced in one tone or another [7–9]. Since the meaning of the word depends on the tone of the sound, new words must be learned along with the tones. There are four tones in Mandarin (the fifth tone is conditional and has no special voice coloring): the first tone is high, even, its melody gives the impression of an unfinished statement; the second tone is an average ascending one, it gives the impression of a second question; the third tone is descending-ascending, gives the impression of a bewildered question; the fourth tone is a high descending one, giving the impression of a categorical command.

2. Materials and Methods

Recall that sound is a mechanical vibration that propagates in the form of elastic waves in various media (gaseous, solid or liquid) [1, 2, 10, 11]. In a narrower sense, sound refers to mechanical vibrations perceived by the senses, in particular the ear (see the Appendix for more details). Among the audible sounds, phonetic sounds, speech sounds, phonemes (which make up speech) and musical sounds (which make up music) are distinguished [10–13].

Sounds can be divided into *tones (simple, complex), noises* and *sonic booms*. A *simple (pure) tone* is a sound that has only one frequency f; it is described by a harmonic oscillation. *The acoustic spectrum* $A(f)$ *of a tone* is the totality of all its frequencies f with an indication of their amplitudes A or intensities $I = |A|^2$. *The main tone* corresponds to the largest amplitude of the spectrum $A(f)$. It is this tone that is perceived by the ear as the *pitch of the sound*. Overtones create the "color" of the sound. Sounds of the same pitch, created by different instruments, are perceived differently by the ear precisely because of the different ratio between the amplitudes of the overtones. *A complex tone* is a sound that contains several frequencies; it is described by a non-harmonic oscillation.

Musical sounds are an example of a complex tone; they contain not one, but several tones, and sometimes noise components in a wide range of frequencies. The acoustic spectrum $A(f)$ of a complex tone is lined; it contains a set of multiple frequencies f [1, 10]. *Noise* is a set of randomly (randomly) changing complex tones of any frequency. The noise spectrum is continuous. *A sonic boom* is a brief sonic impact, such as a bang.

The line spectrum $A(f)$ in the form of a set of individual harmonic components with multiple frequencies f is inherent in musical sounds. In this case, the fundamental frequency determines the pitch of the sound perceived by the ear, and the set of harmonic components determines the timbre of the sound. We also note that there are so-called *formants* in the sound spectrum $A(f)$; the *formants* are stable groups of frequency f components corresponding to certain phonetic elements [10–12]. Formant denotes a certain frequency region in which, due to resonance [1, 14], a certain number of harmonics of the tone produced by the vocal cords are amplified [11–14].

In the sound spectrum $A(f)$, the formant is a fairly distinct region of enhanced frequencies. Formants are denoted by the letter "F". Four formants are usually distinguished to characterize speech sounds (FI, FII, FIII, FIV), which are numbered in ascending order of their frequency: the formant with the lowest frequency is FI, then FII, etc. [2, 10–14]. Formant frequencies are denoted as follows: f_0, f_1, f_2, f_3, f_4 (sometimes use capital letters and start numbering with "1"). For different speech sounds, certain frequency ranges of formants are characteristic. The set of formant values is called the F-picture. Usually, the first two formants are sufficient to distinguish between vowels, but the number of formants in the sound spectrum is always more than two. This indicates more complex relationships between articulation and the acoustic characteristics of sound than if only formants were taken into account: FI, FII. Formants are visualized using spectrograms obtained using digital audio signal processing programs [2, 10, 11, 15].

The formant structure of a particular sound is determined by the characteristics of the formants, i.e. those areas of energy concentration in the acoustic spectrum that are associated with the characteristics of articulation and are necessary for the correct identification of a given sound. The number of formants essential for characterizing speech sounds is defined in different ways. The most common point of view is that four formants are sufficient to characterize the sound, while the first and second formants (FI, FII) are more important than the third and fourth (FIII, FIV).

Two classical theories explain the formation of formants as follows [10]. According to the first theory (Krantzenstein, Helmholtz), the characteristic timbre of a vowel is formed due to the amplification in the supraglottic cavities of one or several harmonic overtones, which arise together with the fundamental tone of the voice in the larynx as a result of complex vibrations of the vocal cords. Due to the vibrations of the vocal cords in the larynx, the fundamental tone of the voice arises together with harmonic overtones. In this case, the supraglottic cavities act as resonators in which amplification of certain overtones occurs, which determines the timbre of the vowel.

The second theory (Hermann) explains the formation of vowel formants as a result of the superposition of the proper tone of the supraglottic cavities, which arises from blowing a stream of air through them, onto the fundamental tone coming from the larynx. Here the formants arise in the oral cavity and are not in a harmonic relationship to the fundamental tone of the voice.

It is important to emphasize the recognition by both theories of the fact that vowel formants are determined by the position of the organs of pronunciation in the supraglottic cavities, and not in the larynx. In this sense, the two theories do not contradict each other.

Earlier in our works, we proposed a number of original approaches for the study of ancient hieroglyphs [8, 9, 16]. In particular, a comparison was made of typical hieroglyphs (keys or Chinese radicals) from modern text messages with hieroglyphs on typical Jiaguwen (Jiǎgǔwén / 甲骨文 are inscriptions on tortoise shells and fortune-telling bones that date back to the 14th–11th centuries BC) and Jinwen (Jīnwén / 金文 are ancient inscriptions on bronze vessels, around the 2nd-1st millennium BC) [3–9, 16, 17]. As a result, matching hieroglyphs and keys were found. An analysis of the most ancient written signs and texts showed that both the hieroglyphs themselves and the structure of the archaic sentence have much in common with the sentence structure of the modern Chinese language. In sum, through the prism of the modification of writing, the historical link between the past and the present day of Chinese civilization was presented.

The articles [9, 16] explored a number of typical hieroglyphs (keys / graphemes / Chinese radicals), such as, 土 (tu) earth, soil; 天 (tian) sky; 卜 (bu) fortuneteller, guess, divination; 册 (ce) letter, message, writing boards; 京 (jing) capital; 宫 (gong) castle; 家 (jia) home, family; 立 (li) stand; 交 (jiao) exchange, transfer, give; and so on. We emphasize that the research methods described in these works are also applicable to other hieroglyphs.

The article [9] examined the history and evolution of the Chinese language. We considered it right to talk about the written language of the Chinese language as the main link between the ancient

Chinese, Middle Chinese and modern Chinese languages. To study the hieroglyphic inscriptions on the ancient Jiaguwen and Jinwen artifacts, some typical hieroglyphs from text messages were compared with hieroglyphs depicted on a bone and a copper vessel. In that order 14 and 16 ancient symbols (hieroglyphs / Chinese radicals) were identified, which are also found in modern texts. At the same time, the possibility of developing methodological foundations for the selection of quantitative criteria is shown, which can be a good addition to traditional methods of studying prehistoric artifacts.

For a comprehensive perspective study, it is also of great interest to study the sound properties of hieroglyphs [18–20], for example, those present on typical samples of Jiaguwen and Jinwen described in our previous works (see [9, 16]). Various models can be used to study the sound spectrum, in particular, models of the following oscillations: a string, a thin plate, a pendulum, a plate in a resonator (see, for example, [1, 2, 10, 11, 13, 14]).

The object of research is the most ancient hieroglyphs (Chinese radicals) included in the hieroglyphic inscriptions on ancient artifacts. At the same time, the main goal (subject of research) is the study of the sound characteristics and parameters of hieroglyphs, i.e. features of their pronunciation, taking into account the syllabic structure (initials and finals) [3, 11–14]. In the future, for simplicity, we will talk about the pronunciation of hieroglyphs (in accordance with the phonetic transcription of Pinyin adopted in China). Note that the pronunciation of hieroglyphs in antiquity (more than 3000 years ago) is not known (see Appendix), so their modern sounding was studied in the work. The main criterion for the selection of hieroglyphs is the simultaneous presence, both in ancient inscriptions and in modern messages.

In our work, to study the characteristics and parameters of the pronunciation of hieroglyphs, we used methods and programs that are used in phonetics for the spectral analysis of sounds with their simultaneous visualization [15]. This eventually made it possible to identify formants in the sound spectrum. Formants are visualized using spectrograms obtained using specialized instruments or computer programs (e.g. “Spectrum Lab”, “TrueRTA (Real Time Audio Spectrum Analyzer)” for Windows) for digital processing of audio signals. These programs are essentially digital appliances capable of replacing specialized instruments as an audio spectrum analyzer. For this purpose, it is possible to use: computer, laptop, tablet or smartphone. The spectrum analyzer shows in real time the frequency spectrum of the analyzed sounds, which can be both audible and inaudible. As a result, the screen shows a spectrogram of the sound of the studied character in a fairly wide frequency range (usually from about 1 to 8000 Hz; there are also applications that have twice the frequency range).

Recall that *visualization* (from Latin *visualis*) is the creation of conditions for visual observation. In a general sense, this is a method of presenting information in the form of an optical image. *Sound visualization* includes methods for obtaining a visible picture of the distribution of certain quantities that characterize the sound (sound field).

The *notation (sheet music)* should also be noted, that is a traditional centuries-old way of visualizing music (musical sounds). This is a very harmonious and concise system, clear and symbolic. Musical notation uses a system of musical clefs and different ways of showing tones and their pitches. From this point of view, there is an analogy of music with the Chinese language.

We have analyzed several models that allow us to explain to a greater or lesser extent the results observed in the experiments. In particular, we revealed the possibility of using some equivalent simplified mathematical models (systems) that allow us to describe the corresponding processes. Within the framework of these different types of resonance phenomena and factors influencing them were also considered within the framework of these models. The conducted analysis has shown that there is a possibility to choose different fundamental frequencies (tones) of oscillations of the investigated equivalent systems. At this stage of research we have chosen as a basic model in which the observed speech spectrum is determined both by the complex oscillations of the vocal cords in the larynx and by the position of the pronunciation organs in the supraglottic cavities.

3. Results and discussion

To demonstrate the capabilities of the research method described in this article from previously studied hieroglyphic signs [8, 9, 16] as an example, the following three hieroglyphs (Chinese radicals) were chosen: 家 "home, family"; 立 "stand" and 交 "exchange, transfer, give". It can be noted that many hieroglyphs carry a certain historical context and symbolism that is still understood. For example, the upper part of the character house 家 means a roof, and the lower part means a pig. In ancient China, a pig under a roof (that is, in a house) was a sign of the wealth and success of a given family or clan.

To describe the spectrograms $A(f)$ of the pronunciation of hieroglyphs (keys / Chinese radicals), the pendulum model was used in the work (for more details, see the Appendix). This model is considered to be simplified, but it allows one to describe and study the spectra of sound vibrations, i.e. the resulting spectrograms $A(f)$ of the pronunciation of hieroglyphs [18–20]. We will consider the oscillation frequencies f found using the pendulum model as some parameters that allow us to characterize the pronunciation of hieroglyphs (see Appendix).

To receive digital audio files, pronunciation of hieroglyphs 家, 立 and 交 several sources were used, in particular: 1) audio files from open sources (see, for example, [18–20]); 2) digital software application "Trainchinese"; 3) professional Chinese translator. This article presents only the latest spectrograms; however, other sources were also used in the analysis of the obtained data. Fig. 1 shows successively some of the obtained sound spectra $A(f)$ (dependence of the amplitude A of sound vibrations on frequency f, i.e. spectrograms) of three studied hieroglyphs. Let's compare the found frequencies of model vibrations (see Appendix) with the spectrograms of the pronunciation of hieroglyphs corresponding to the selected hieroglyphs: 家 "home, family", 立 "stand" and 交 "exchange, transfer, give".

Recall that a spectrogram $A(f)$ is a visual representation of the frequency spectrum of a signal that changes with time t or frequency f [1, 2, 10, 11, 13]. When applied to an audio signal, spectrograms are sometimes referred to as sonographs, voice prints, or voice charts. Spectrograms are widely used, for example, in such areas as music, linguistics, speech processing, sonar, seismology, and others.

Any process (oscillation) $A(t)$ that is non-periodic in time can be represented as an infinite sum (the integral is taken over infinite limits) of oscillatory processes (oscillations) periodic in time:

$$A(t) = \frac{1}{2\pi} \int A(\omega) \exp(i\omega t) d\omega, \tag{1}$$

where $\pi \approx 3.14$; $A(\omega)$ (or $A(f)$) is the spectrum (spectrogram) or spectral amplitude density, $A(f)$ is the acoustic characteristic; in the quasi linear approximation $A(f) = E(f)K(f)$, where $E(f)$ is an input acoustic signal (its spectral amplitude density), and $K(f)$ is the transmission coefficient of the vocal tract (acoustic characteristic of a filtering process); $\omega = 2\pi f$ is the angular frequency of process (oscillation), f is the frequency measured in hertz (Hz). The expression (1) is usually called the Fourier integral of the oscillatory process. In practical applications, expansion $A(t)$ into a finite series (the sum of a finite number $M \ll \infty$ of harmonics) is used:

$$A(t) = A_0 + \sum_{m=1}^{M} A_M \cos(\omega m t) + B_m \sin(\omega m t), \tag{2}$$

where A_0, A_m and B_m are well known Fourier coefficients.

A spectrogram (see Eq. (1)) can be obtained using: a spectrometer, a band pass filter, a Fourier transform, or a wavelet transform (see e.g., [1, 2, 10–12]). The most common way to represent a spectrogram is a graph with two dimensions: one axis (horizontal) represents frequency f (or time t) and the other axis (vertical) represents the amplitude A of the sound for a particular time or frequency value (see Fig. 1). As a result, a spectrogram $A(f)$ of sounds is visible on the display screen. The spectrograms $A(f)$ of the sounds of the three hieroglyphs under study in the frequency f range from around 1 to 2000 Hz are shown in Fig. 1 (the vertical axis is the normalized level (loudness) of the sound A, and the horizontal axis is the frequency f in Hz).

A comparative analysis of the obtained data showed that in the initial interval (approximately from 1 up 800 Hz) all spectrograms $A(f)$ contain sufficiently intense low frequencies f, which contain a range from the fundamental frequency to a certain boundary value of about 1–5 Hz and which naturally have a low-pitched sound. The *fundamental tone* in our simple (basic) model has a *fundamental frequency* f_0 about 100 Hz, and the overtone frequencies are multiples of the fundamental tone (see the Appendix for more details). Moreover, in all spectra $A(f)$ there is a section in the frequency f range from around 1 Hz up about 25–30 Hz with characteristic maxima in the vicinity of about 10–20 Hz. Any oscillatory processes are accompanied by the generation of infrasonic waves. The most important source of infrasound for us is the process of speech production. This is precisely what spectrograms demonstrate.

As can be seen from the spectrograms $A(f)$, individual characteristic discrete components stand out against the background of continuous sound spectra (the line part of the spectrum). Vibrations that have a line spectrum give the impression of a sound with a more or less defined pitch. Such a sound is called *tonal*. On the spectrograms of the pronunciation of hieroglyphs, line spectral components stand out; this indicates the tonality of sounds.

The *pitch* of the tonal sound is determined by the fundamental (lowest) frequency. Oscillations with frequencies f that are multiples of fundamental frequency f_0 are *overtones*. The ratio of the intensities of the main tone and overtones determine the *timbre of the sound*; give sound a certain "color". The phases of harmonics do not affect the timbre of the sound. In the absence of overtones, a tonal sound is called a *pure tone*. A pure tone is given by tuning forks, which are used when tuning musical instruments.

In the study, one should also take into account that in the process of the formation of one Chinese language, there were many different dialects, where the same hieroglyph could be pronounced differently. This fact allows speaking about the possibility of using some extensive discrete set of sounds (quasi-continuum, i.e. a sufficiently large but finite set of close but not identical pronunciations) for each specific character (see comparison with Standard Chinese in Appendix).

The spectrogram $A(f)$ of the hieroglyph 家 is shown in Fig. 1(a). A study of the spectrogram $A(f)$ of this hieroglyphic sign showed that it contains at least 6 overtones corresponding to the fundamental frequency f_0. We emphasize that in the spectrogram $A(f)$ of the pronunciation of the hieroglyph "home, family" there are three characteristic local maxima in the vicinity of frequencies f about 300, 600 and 1000 Hz. The main audible frequency f (about 15–20 Hz) determines the audible pitch of the hieroglyph 家, and the set of harmonic components is the timbre of this sound.

The spectrograms $A(f)$ of other hieroglyphs 立 and 交 (see Fig. 1) reveal similar features. Note that the maximum in the vicinity of 300 Hz is the main one in the spectrum; the amplitude $A(f)$ of the sound wave in the vicinity of this frequency f is greater than all other amplitudes of sound waves in the spectrogram of the hieroglyph "stand" 立.

At this stage of research, it is possible to propose as the simplest sound model for these hieroglyphic characters the basic model of hieroglyphic pronunciation, in which the spectrum $A(f)$ is limited from above by a frequency f, for example, close to 1200 Hz. Then their spectrograms $A(f)$ in the range from about 1 up 1200 Hz will have 3 spectral components f (formants F), which are overtones for the

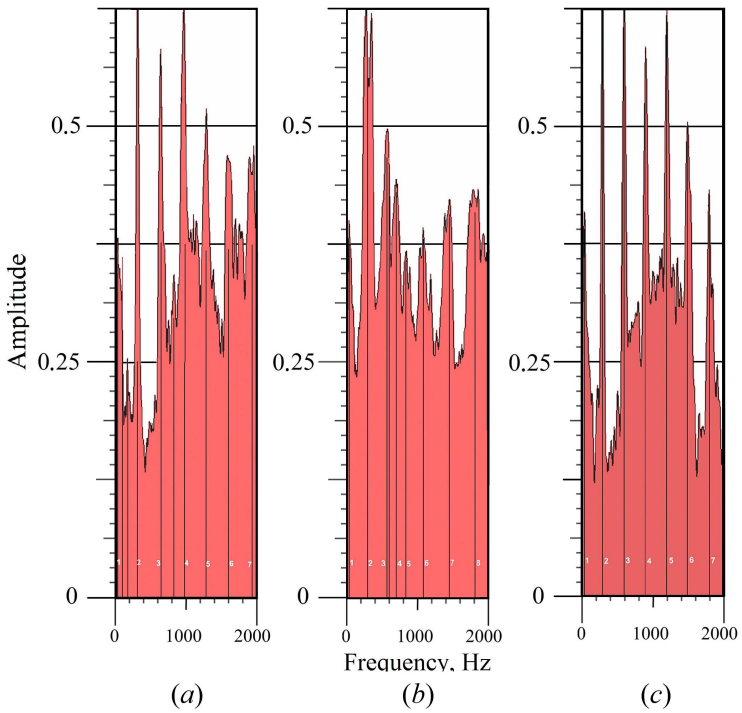

Figure 1. Spectrograms $A(f)$ of the pronunciation of hieroglyphs in the range from 0 to 2000 Hz: (a) 家; (b) 立; (c) 交

fundamental frequency $f \approx 100$ Hz and one ($f \approx 10$ or $f \approx 30$ Hz) that corresponds to the inverse bifurcation. These low frequency components correspond to sub harmonics of the following type: $f_0/10$ or $f_0/3$. It should be noted that the last two cases correspond to an increase in the oscillation period by 10 and 3 times, respectively. Since the oscillation period depends directly proportionally on the resonator size [1, 14], we can conclude that the (effective) resonator size increases dynamically during pronunciation. In simple terms, we can assume that the fundamental tone of the voice (along with harmonic overtones) arises in the larynx due to the oscillations of the vocal cords, and then in the supraglottic cavities, which increase the size of the effective resonator, some overtones are amplified. As can be seen from Fig. 1, discrete harmonic components are observed in the spectra against the background of continuous spectral bands. This indicates the presence of quasi-harmonic and non-periodic (noise) components in the original spectrum. It can be seen that subharmonics near the fundamental frequency have the highest intensity. It can be assumed that the appearance of non-periodic oscillations of different components of the speech tract up to a certain point contributes to the generation of subharmonic components. A more detailed analysis of the phenomenon of the appearance of subharmonics will be considered in one of the subsequent works.

We will name this simple model including overtones and subharmonics a fundamental (basic) "F-model" of the pronunciation of hieroglyphs in accordance with the phonetic transcription of Pinyin adopted in China. Taking a subharmonic component into account highlights the difference between our model and others, and also demonstrates the novelty of our scientific approach and its relevance.

Within the framework of this model, the hieroglyphic signs under consideration will have the following four bands containing four main approximate frequency maxima in spectrograms $A(f)$

(Hz): 1) 家 “home, family”: 10, 300, 600, 1000; 2) 立 “stand”: 30, 300, 600, 1000; 3) 交 “exchange, transfer, give”: 30, 300, 600, 900 (or more exactly: a frequency band near 850–1050 Hz). At this stage, the possibility of distinguishing sounds will remain if they are synthesized from these frequencies within the framework of this model. Indeed, a sequential comparison of two of the three spectrograms $A(f)$ shows that there is at least one non-coinciding spectral component, which will make it possible to distinguish these sounds.

This analysis suggests that the model vibration frequencies found by us can correspond to at least some (low-frequency) part in the spectrograms $A(f)$ of the original (dialect) sounds of the 3 studied hieroglyphs: 家, 立 and 交. At the same time, a correspondence was found between the overtones of the found frequencies of the model oscillations and some parts of the spectrograms of these hieroglyphs.

Using the methods of harmonic analysis, the spectra $A(f)$ were numerically synthesized, which made it possible to approximately describe the found values of the formants of the studied hieroglyphs. To do this, we used the values of the fundamental frequency of about 100 Hz and its overtones and subharmonics in the frequency range from about 10 to 1200 Hz. The study took into account that the spectrograms of the corresponding hieroglyphic signs 家, 立 and 交 in a given range have accordingly 4 spectral frequency components f (Hz): 10 (or more precisely: a band around 5–20 Hz), 300, 600, 1000; 30, 300, 600, 1000; 30, 300, 600, 900. To control the accuracy, audiograms in the time domain were also used dynamic spectrograms $A(t)$ (see Appendix). In particular, the time taken for the decrease in the volume of experimental and calculated sounds to zero was taken into account. In this parameter, the experimental and calculated data are in good agreement with each other.

In the future, this procedure can be extended, for example, to a frequency of 2000 Hz or more. However, it is clear that the expansion of the spectral range will complicate the solution of this problem. For a more accurate synthesis of model spectra $A(f)$, in addition to the main harmonics (tones), one can add a certain set of auxiliary spectral components such as overtones and subharmonics or noise-like sound components (we will speak next only about spectral components without distinguishing between them). It should be noted that such signals are approximate models of the studied sounds.

So, at this stage, some basic principles for constructing formant models for the hieroglyphic signs “home, family”, “stand” and “exchange, transfer, give” in the selected frequency range are formulated. Such a construction can be based, for example, on the principle of specifying a certain frequency range in the studied spectrograms, taking into account the presence of certain spectral components in them. At this point, one should rely on the form of the original acoustic spectrograms.

We have shown that these three groups of frequencies f (spectral components 10, 300, 600, 1000; 30, 300, 600, 1000; 30, 300, 600, 900 Hz) can be considered as some formants F of the first (basic) model of the pronunciation of the hieroglyphs “house, family”, “stand” and “exchange, transfer, give” in the frequency range from about 10 to 1200 Hz. These stable groups of frequency components correspond to a simple model of phonetic sounds for given hieroglyphs: 家, 立 и 交. As a result we can construct a simplified basic pronunciation “F-model”, containing two FI, FII or four formants FS, FI, FII, FIII (FS denotes a subharmonic formant). We will call this model a basic “F-model” of the pronunciation of hieroglyphs in accordance with the phonetic transcription of Pinyin adopted in China. A further simplification of this model is to use the concept of resonant bands (spectral bands concentrated near resonant frequencies) and their weight contribution to the spectrum, taking into account their amplitude. As can be seen from Fig. 1, from this point of view the components below 10 Hz can be neglected at the initial analysis. Therefore in this case our simplified pronunciation “F-model” can include: 1) two formants (FI, FII), contained in frequency bands around approximately 300, and 600 Hz respectively; 2) or four formants (FS, FI, FII, FIII): 20 (average for two subharmonics; there formant is denoted as “FS”), 300, 600, and 1000 Hz.

The data obtained by us were compared with the spectrograms of the pronunciation of hieroglyphs in Mandarin (standard) dialect of the Chinese language [18–20]. It is established that they correspond to each other. In addition, our results were compared with some data from the paper [11].

The results obtained by us testify to the correctness of the assumptions made and chosen research methods, which made it possible to propose formant models for hieroglyphs that are found simultaneously in ancient inscriptions and in modern texts. So an idea was obtained not only about the place and role of formants in the spectrograms of the Chinese language, but also about their certain universality.

Within the framework of the described model, it is possible to create a library of basic pronunciations of hieroglyphs for a typical 100-character key table of Chinese characters (an example of some conditional analogy with a 100-word Swadesh list). Such a music library will be useful, for example, for automatic preliminary recognition of digitized spectrograms of the pronunciation of hieroglyphs depicted on artifacts. Experts can work at the next stage, carrying out a qualified systematization of the studied ancient hieroglyphic inscriptions.

In conclusion, it is worth noting several areas in which various methods of audiovisual research are used, demonstrating their purpose, features, capabilities and differences (see, for example, [15, 21, 22]). The growing interest in this area is partly due to the recently actively developing trend of visualization of various Internet resources, as well as the increasingly widespread use of advanced innovative technologies (Artificial Intelligence (AI) and Virtual Reality (VR)). Due to the large number of publications on these topics, we will consider only three works as examples [15, 21, 22] (see the list of references for this article, as well as the literature cited in these publications [15, 21, 22]).

One of these areas is related to television [15]. Here, audiovisual analysis aims primarily at understanding the ways in which a sequence (or an entire film) uses a combination of sound and image. At the same time, it is proposed to give answers to the questions: “What do I see?”, “What do I hear?” Audiovisual analysis must be word-based at its core and attempt to develop ways of classifying sound using new descriptive criteria. Thus, audiovisual analysis is considered in the work [15] as a descriptive analysis which should avoid any interpretation of a psychoanalytic, psychological, social or political nature. However, the interpretation can follow based on the results of the analysis done. An example of an approach to audiovisual analysis of images of water and waves is given. It is noted that the analysis is not interested in the symbolism of water and waves, but in the wave as a dynamic model [15]. From our point of view, this approach is a systems approach, when a complex phenomenon is studied as an integral system. A similar approach is quite applicable to theatrical video productions, as well as to traditional theater, Pantomime Theater, theater of facial expressions and gestures, and shadow play when sounds may be present in a minimal amount or absent altogether. In the latter case, the absence of sounds is replenished or supplemented by each viewer individually in accordance with his life experience, individual skills, education, knowledge of languages and cultural traditions.

The article [21] analyzes subjective experiments examining the relationship between audio and video quality measured separately and the overall quality of the audiovisual experience. At the same time, the paper provides a statistical analysis of this experiment within the framework of model approaches. From the analysis of the data obtained, we come to the conclusion that there is no obvious natural relationship between the accuracy of the multiplicative model and the authors’ conclusions as the dominant factor (sound quality or video quality) on the overall audiovisual quality. In the multiplicative model, people combine audio and video errors together using the multiplicative rule. And the true formula that describes the statistical data depends on the context and the test material in question. All of this indicates that audio quality and video quality are equally important to the overall audiovisual quality. In particular, one important factor that can affect audiovisual quality is noted: audiovisual synchronization errors (for example, lip synchronization).

The use of audiovisual analysis methods for the purpose of selecting and forming the necessary audiovisual content can be also promising in a variety of educational processes. The effective use of such an approach, for example, in the creation of audiovisual media presented as new perspective educational platforms, can contribute to the formation of the necessary professional competencies, for example, in the training of specialists in the field of journalism, public relations and advertising [22].

We presented in this paper a fundamental (basic) "F-model" of the pronunciation of hieroglyphs in accordance with the phonetic transcription of Pinyin adopted in China. This model includes overtones and subharmonics. Taking a subharmonic component into account highlights the difference between our model and others, and also demonstrates the novelty of our approach, its relevance and the prospect of its development in further studies.

Appendix

Let us give brief information about the sound (sound waves). Sound waves are an example of an oscillatory process [1, 2, 10]. The simplest sound wave is a periodic (i.e., the amplitude values are repeated at regular intervals) oscillation described by a sinusoid. The sound is characterized by a number of parameters. Here are some of these parameters: amplitude, intensity, wavelength, oscillation period, oscillation frequency. Only the last two parameters are important to us for research: the period of oscillations is the smallest time interval for the repetition of oscillations (time t is measured in seconds (s)); the oscillation frequency f is inversely proportional to the period T (the unit of frequency is Hertz (Hz)). The human ear perceives sound frequencies f in the range of approximately 15 up 20000 Hz. A more detailed description of sound waves is beyond the scope of the article (see details in [1, 2, 10, 11, 13, 21]).

Let us recall the model of an ideal pendulum. In this model, the load is suspended on a thread and performs harmonic oscillations about the equilibrium position (see below Eq. (4) and accompanying explanations). In this case, the deviation angle is small, the mass of the load is greater than the mass of the thread, and the length of the thread is greater than the dimensions of the load. The period of oscillation in this mathematical model can be found as:

$$T = 2\pi\sqrt{l/g}, \tag{3}$$

where l is the length of the pendulum; g is the gravity acceleration ($g \approx 10$ м/с^2); $T = 2\pi/\omega_0$, so $\omega_0 = \sqrt{g/l}$, $\omega_0 = 2\pi f_0$, f_0 is the own frequency of the system. The pendulum can be considered as a kind of mechanical resonator.

This model describes the vibrations of a tuning fork characterized by a pure tone. That model will be closer to the real system it describes, in which the oscillations are complex, if the action of some external driving force is introduced into it. This will result in more frequency components than just one fundamental tone. In our case, this may be due to some influence on the pendulum (located, for example, in a cavity resonator): pushing the swinging weight, changing the length of the thread. If the impact, for example, a light push on the load on the thread, is in time with the oscillations, then both the fundamental frequency of oscillations and the occurrence of some resonances in this system are possible [1, 2, 10, 14]. In the resulting overtones (or harmonics), the frequency is an integer number of times higher than the frequency of the fundamental tone, and the intensity is weaker, the higher the frequency. The human vocal cords make complex sounds. The presence of complex vibration movements, which form both the main tone and overtones, is one of the reasons for the emergence of complex sounds.

We will consider the pendulum as some simplified (basic F-model) that allows us to describe the main properties of the spectrograms of the pronunciation of hieroglyphs. Indeed, the oscillations are nearly harmonic only at very small angles. At large angles oscillations turn into anharmonic vibrations. As a result, oscillations with frequencies 2ω, 3ω , etc., appear where the fundamental frequency of the oscillator is ω. Moreover, the frequency ω deviates from the frequency ω_0 of the harmonic oscillations. As the first approximation, the frequency shift $\Delta\omega = \omega - \omega_0$ is proportional to the square of the oscillation amplitude: $\Delta\omega \propto A^2$. In a system of oscillators with different natural frequencies anharmonicity results in additional oscillations with combined frequencies and one can observe, for example, intermodulation and combination tones.

Consequently, we replace the complex movements of the human tongue in the process of pronunciation with simpler pendulum oscillations. We will call appropriate model like basic "F-model" of the pronunciation of hieroglyphs (in accordance with the phonetic transcription of Pinyin adopted in China). Obviously, this is a fairly simplified mathematical model of the pronunciation of hieroglyphs. However this model allows us to find the frequency of fundamental vibrations and frequencies of overtones, i.e. desired sound spectrum $A(\omega)$. And this allows finally explaining the identified features of F-pictures of hieroglyphs.

The tongue is one of the main articulators in the production of speech. For example, different vowels are pronounced by changing the pitch of the tongue and retracting it to change the resonant properties of the vocal tract. These resonant properties enhance certain harmonic frequencies (formants), which are different for each vowel, and attenuate other harmonics. Consonants are articulated by constricting the flow of air passing through the vocal tract. Then, many consonants have a narrowing between the tongue and some other part of the vocal tract.

The solution of such problems, even in the first approximation, encounters serious theoretical and computational difficulties. At the same time, studies of such systems are undoubtedly relevant and promising, since they make it possible to describe the behavior of a number of biological and social processes. We only note that the main problem is connected with the fact that there is no general theory of oscillations of strongly nonlinear systems. To study the model of a pendulum under the action of a quasi-periodic driving force, one can use, for example, a theory describing the interaction of two coupled (quasi-)linear oscillators. It is important to note that even in the absence of resonances; there is instability in the behavior of this system. If we introduce into the model some slight "jitter" of the axis of rotation of the pendulum (relative to the suspension point), then we can make the behavior of the system more stable. In non-resonant (rather far from resonance regions of the studied structure) and resonant (near resonance regions) cases, asymptotic stability is possible in such a system.

To illustrate, we can cite the case when the external driving force changes according to an (almost) harmonic law. Then the oscillations are described by the following second order differential equation:

$$\frac{d^2A}{dt^2} + 2\beta\frac{dA}{dt} + \omega_0^2 A = B_0\cos(\omega t), \tag{4}$$

where β is the damping coefficient, ω_0 is the own frequency of the system, B_0 is the amplitude of the driving force, ω is the frequency of the driving force.

The dependence of the amplitude A of forced oscillations on the frequency ω of the forcing force leads to the fact that at some frequency determined for a given system (4), the amplitude of oscillations reaches its maximum value.

The vibrating system is particularly responsive to the action of the forcing force at this frequency. This phenomenon is called *resonance*, and the corresponding frequency is called the *resonant frequency*. The value of the resonance frequency ω_r is equal to the value:

$$\omega_r = \left(\omega_0^2 - 2\beta^2\right)^{1/2}. \tag{5}$$

It follows from the obtained expression (5) that, in the absence of medium resistance, the amplitude of oscillations at resonance theoretically tends to infinity. According to (5), the resonance frequency at condition $\beta = 0$ coincides with the own frequency ω_0 of oscillations of the system. The dependence of the amplitude $A(\omega)$ of forced oscillations on the frequency of the forcing force (or on the frequency of oscillations) has a well-known form of resonance curves similar in shape to a Gaussian curve.

We emphasize that the smaller β is, the higher and to the right lies the maximum of the resonance curve. At large damping (when $2\beta^2 > \omega_0^2$) the expression for the resonance frequency becomes imaginary. This means that in this case resonance is not observed in the system. With increasing frequency, the amplitude of forced oscillations monotonically decreases, i.e., asymptotic stability is observed in the system.

Another type of influence on the system consists in the (almost) periodic change of some parameter of the system, for example, the length of the pendulum, in time with the oscillations. In this case, resonance is also possible, which is called *parametric resonance*. For this purpose, it is necessary, for example, to periodically change the length of the pendulum, increasing it at the moments when the pendulum is in extreme positions, and decreasing it at the moments when the pendulum is in the middle position. As a consequence, the pendulum will swing wildly. By controlling the moments of influence on the pendulum length, one can also provide asymptotic stability.

To reveal the effect of asymptotic stability in spectrograms, experimental frequency (amplitude-frequency) and dynamic (time) spectrograms were studied. As an example, one can see Fig. 2 (the vertical axis is the frequency f in Hz, and the horizontal axis is the time t in seconds). The growth of the sound amplitude in Fig. 2 is shown by the darkening of the spectrogram $A(f)$. For comparison, using harmonic analysis methods, the spectra $A(f)$ of sound signals (having three main spectral components f in the range from about 1 up 700 Hz), similar to the experimental ones, were numerically synthesized.

Then, using the inverse Fourier transform of the data of the amplitude-frequency spectra $A(f)$, the corresponding audiograms in the time domain were obtained: dynamic spectrograms $A(t)$. Their analysis showed that the asymptotic stability of the studied sounds is indeed observed, which manifests itself in the gradual decay of the amplitude $A(t)$ of the sound signals with time (see Fig. 3, where the vertical axis is the normalized amplitude of the sound, and the horizontal axis is the time in seconds). It can be seen that the decrease in the calculated amplitude occurs rather quickly, approximately in 0.2 s. This decay process in time can be described by the normalized exponential decreasing dependence with an attenuation coefficient about 50:

$$A(t) \propto A_0 \exp(-50t), \tag{6}$$

where A_0 is the initial value of the amplitude $A(t)$, e.g. for $t = 0$.

It is important to emphasize that the effect of the asymptotic stability of dynamic spectrograms $A(t)$ manifests itself not only in the temporal, but also in the energy localization of the corresponding sounds. This effect causes the volume of sounds to fade gradually over a period of approximately 0.02–0.05 s.

The effect mentioned above is based on both the actual short-term nature of the pronunciation of hieroglyphs (localization of the process in time) and the properties of the human vocal tract (localization of the process in space). Thus, in the characteristics (spectrograms) and parameters (formants, time intervals) of the pronunciation of hieroglyphs, their syllabic structure (initials and finals) was naturally reflected. From a theoretical point of view, this two-component "initial-final" structure can be described as the interaction of two connected oscillators, each of which is responsible for its own zone in the human cerebral cortex. In this case, one oscillator can be assigned a zone that causes excitation, and the second — a zone that causes inhibition. As a consequence, there

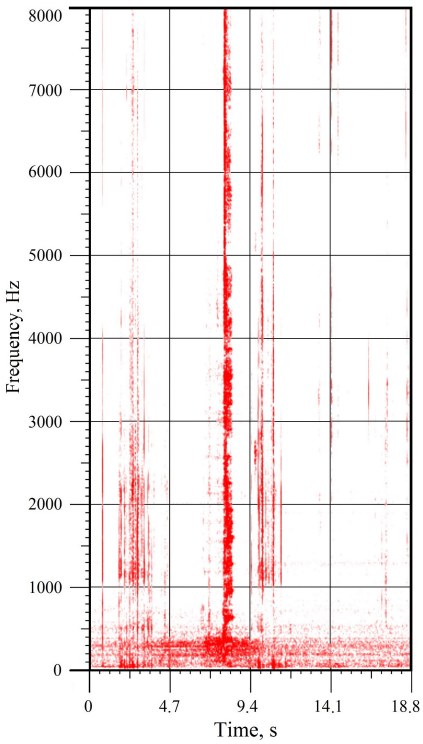

Figure 2. Dynamic spectrogram $A(t)$ of the hieroglyph "home, family" 家

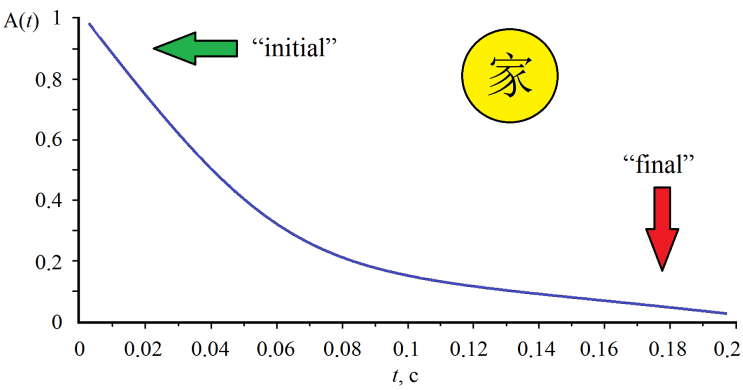

Figure 3. Estimated dynamic spectrogram $A(t)$ of the hieroglyph 家

is a certain balance of these processes and the final asymptotic stability, which is shown in Fig. 3. A more detailed analysis of this phenomenon is beyond the scope of this article (see e.g. [1, 2, 11, 13, 14]).

Note that the decrease in the calculated amplitudes for dynamic spectrograms $A(t)$ almost to zero occurs approximately in 0.2–0.7 s (see Eq. (6) and Fig. 3). Let us compare these data with the

experimental data shown in Fig. 2. Dynamic spectrogram $A(t)$ of the actual pronunciation of the hieroglyph 家 observed over a period of time from 7.7 to 8.2 seconds, i.e. about half a second and corresponds to the time predicted by calculations (similar data were also obtained for a number of other hieroglyphs). We emphasize that the darker region on the dynamic spectrogram $A(t)$ (see Fig. 3, center of the time interval approximately from 7.5 to 8.5 seconds in the frequency band from 1 to 8000 Hz) correspond to a greater sound volume (on Fig. 4 is an interval from approximately 0 to 0.1 s).

As a consequence, for the study of such complex systems, preference is given to approximate qualitative and numerical analysis. In this case, the possibility of reducing such quasi linear systems to simpler equivalent systems can be used. So we can replace the complex movements of different components of the human speech tract with simpler oscillations (see Eq. (3) and Eq. (4)). A more detailed description is beyond the scope of this article (see, for example, [1, 2, 10–14, 21]).

From Fig. 1 and Fig. 2 we can see the effects of the sound source (generator) (larynx and vocal cords) on the filtering system of the speech tract (pharynx, oral and nasal cavities). To separate the contributions of the sound source and the speech path, we can take for simplicity in expression (1) (or its corresponding Fourier image in the time domain) E as a delta function type, i.e., a unit pulse with spectral density equal for all frequencies of the speech path. Then the spectral density $A(f)$ of the acoustic signal at the output is equal to $K(f)$, i.e. the transmission coefficient of the speech path (reflecting the filtering property of the delta function).

In this case, the contribution of elements to the characterization of the speech path can be most simply described: $K(f) = k(f)K_1(f)K_2(f)K_3(f)K_4(f)$, where $k(f)$ is the correction factor, $K_{1-4}(f)$ are the resonant frequency responses (in general, complex-valued) of the corresponding formants FI–FIV. As a result we can propose a corresponding basic "F-model" for the pronunciation of hieroglyphs.

Under this assumption, we see in Fig. 1 the actual characteristic $K(f)$ of the speech path fully describing the F-pattern. In our simple base "F-model" the component corresponding to the fundamental frequency near 100 Hz characterizes the sound source (larynx and vocal cords). As a consequence, our base "F-model" describes in this approximation exactly F-patterns of the pronunciation of hieroglyphs. From Fig. 2 for the moment of pronunciation of the hieroglyph 家, observed during a period of time from 7.7 to 8.2 seconds, we can see the vertical profile of $A(f) \approx K(f)$ for any fixed moment in the specified time interval about 0.5 s. Further complication of the base "F-model" can be done in particular by taking into account the formant corresponding to the fundamental frequency.

Now we can give a brief description of our proposed model "F-model" of the studied phenomenon of the pronunciation of hieroglyphs. In addition, we will give some explanations to this model, allowing us to understand the essence of the key idea and its relationship to the phonetic transcription of Pinyin adopted in China. This is a rather simple mathematical model of the pronunciation of hieroglyphs. However this model allows us to find the frequency of fundamental vibrations and frequencies of overtones, i.e. desired sound spectrum $A(f)$ (and $K(f)$). This ultimately allows us to explain the recognized features of F-patterns of hieroglyphs. Using the inverse Fourier transform of the data of the amplitude-frequency spectra $A(f)$, the corresponding audiograms in the time domain can be obtained, i.e. dynamic spectrograms $A(t)$. In useful applications, expansion $A(t)$ into a finite series (the sum of a finite number of harmonics, see Eq. (2)) is used.

This model also includes an investigation of the temporal asymptotic stability of these sound characteristics of hieroglyphs, as it is one of the most important indicators of the phenomenon of the pronunciation of hieroglyphs under study. For this purpose, the study of dynamic dependence $A(t)$ is carried out; in particular, the type of a suitable function approximating the experimental one $A(t)$ under study is found. It is convenient to take an exponential damped function as the latter. The degree of its closeness to the dynamical dependence $A(t)$ under study can be estimated, for example,

in the mean-square metric (i.e. RMS error). In our case, the RMS error of approximation using functions of the form in Eq. (6) for a number of conducted experiments did not exceed 10–20%.

4. Conclusion

In summary, the construction of the base "F-model" boils down to obtaining the following 8 main components. 1) Images of studied hieroglyphs (Chinese radicals) and their digital files; all with detail descriptions. 2) Digital audio files of pronunciation of studied hieroglyphs. 3) Experimental sound spectrum $A(f)$ of the studied hieroglyphs and $K(f)$. Highlighting four basic formants in $A(f)$ (and $K(f)$). 4) Construction based on the experimental amplitude-frequency spectrum $A(f)$ a simplified basic "F-model" pronunciation of hieroglyphs under study, containing two, three or four formants (FI, FII; FI, FII, FIII; or FS, FI, FII, FIII, where FS denotes a subharmonic formant). We will call this models a basic "F-model" of the pronunciation of hieroglyphs in accordance with the phonetic transcription of Pinyin adopted in China. If the components below 300 Hz are neglected, a simplified pronunciation "F-model" will include only three formants FI, FII, FIII, contained in frequency bands around approximately 300, 600 and 1000 Hz respectively. 5) Computation of the inverse Fourier transform of the data of the amplitude-frequency spectra, i.e. dynamic spectrograms $A(t)$. 6) Study of temporal asymptotic stability of dynamic spectrograms $A(t)$ of the studied hieroglyphs. 7) Study of peculiarities of the pronunciation of hieroglyphs (localization of the process of pronunciation in time and localization of the process of the pronunciation in space), enabling the identification of the syllabic structure of hieroglyphs, i.e. initials and finals. Calculation of frequency and time localization parameters characterizing this two-component "initial-final" structure of the pronunciation of hieroglyphs. 8) Creation of a digital databank of the hieroglyphs under study within the framework of this basic "F-model". This databank should include the following elements in digital formats: images of studied hieroglyphs; audio and audio-visual files of pronunciation of studied hieroglyphs; $A(f)$ (and $K(f)$); formants FI, FII, FIII, FIV (or FS); formant frequencies f_{1-4}; $A(t)$; set of suitable functions like $\exp(-at)$ (a is a certain constant), approximating with the required accuracy the experimental data under study; set of localization parameters in frequency and time domain, characterizing two-component "initial-final" structure of the pronunciation of hieroglyphs. It is important to underline that there are no analogous models of the pronunciation of hieroglyphs presented and described in similar formalized and structured way in the scientific literature.

For an introduction to more complex models, including physical models, as well as studying speech production in different specific languages, which attempt to describe the features of a human speech, we recommend, for example, the following publications [23–28]. At the same time, we consider that our "F-model" can also be used in innovative technologies like AI and VR.

Recall that a periodic oscillation can be represented as a sum of harmonic oscillations. A real pendulum has complex oscillations, therefore, in addition to the fundamental tone; oscillations with a higher frequency are also formed. The frequency of these oscillations corresponds to the frequency of sound waves, the spectrum of which also contains frequencies that are multiples of the fundamental tone, i.e., exceeding it by a multiple number of times: 2, 3, 4, etc. (these numbers are the spectrum harmonic number M, see Eq. (2)).

A more detailed examination of the type of the calculated approximate dynamic spectrogram near zero showed that there is a section of increasing amplitude to the maximum shown in Fig. 3. Next, using the Fourier transform of the data of the dynamic spectrogram of hieroglyph 家, the corresponding amplitude-frequency spectra in the frequency domain were obtained (it looks like a Gaussian curve modulated by a damped cosine). Analysis of the obtained data demonstrated that the experimental subharmonic formants about 12 and 30 Hz are slightly less than calculated

model subharmonic formant values: 18 and 36 Hz. This indicates the need for further theoretical, computational and experimental study of the behavior of the audio spectrum in the frequency range below 300 Hz to understand the nature of this frequency shift, especially in the subharmonic region.

Preliminary we put forward a hypothesis that this frequency shift in a spectrum can be partly caused by a little change of a modern pronunciation of hieroglyphs in comparison with their most ancient sounding. This idea made it possible to explain partly and approximately the identified features of F-patterns of hieroglyphs. The essence of the hypothesis is that these features are associated with the process of historical formation of the Chinese people, the most important stage of which is the transition from a sedentary lifestyle to a more mobile, active lifestyle. Probably similar phenomena can be found in F-patterns of other people. To analyze the peculiarities of linguistic information $I(x)$ dissemination in a similar community, one can use, for example, such a mathematical nonlinear model: $I(x) = \lambda_1 x(1-x) + \lambda_2 x(1-x)^2$ (see e.g. [8, 9]). Parameters of this model: x is the independent variable (varies from 0 to 1); $\lambda_{1,2}$ are coefficients (or control parameters) of the analyzed model (see for more details [8]).

Author Contributions: Conceptualization, A.E. and M.E.; methodology, A.E. and M.E.; software, A.E.; validation, M.E. and A.E.; formal analysis, M.E. and A.E.; investigation, A.E. and M.E.; resources, M.E. and A.E.; data curation, M.E. and A.E.; writing–original draft preparation, A.E.; writing–review and editing, M.E. and A.E.; visualization, M.E. and A.E.. All authors have read and agreed to the published version of the manuscript.

Funding: The publication was prepared with the partial support of the RUDN "Strategic Academic Leadership Program".

Data Availability Statement: Data sharing is not applicable.

Acknowledgments: The authors would like to thank their colleagues for their kind feedback, which helped in preparing the paper for publication

Conflicts of Interest: The authors declare no conflict of interest.

Declaration on Generative AI: The authors have not employed any Generative AI tools.

References

1. Rocchesso, D. *Introduction to sound processing* (Phasar Srl, Firenze, 2003).
2. Bondarko, L. V., Verbitskaya, L. A. & Gordina, M. V. *Fundamentals of general phonetics* 4th (Academy, St. Petersburg, 2004).
3. Yakhontov, S. Y. *Ancient Chinese language* (Nauka, Moscow, 1965).
4. Vasiliev, L. S. *Ancient China: in 3 volumes* (Oriental Literature, Moscow, 1995; 2000; 2006).
5. *Atlas of the languages of the world. The origin and development of languages around the world* (Lik press, Moscow, 1998).
6. *The peopling of East Asia: putting together archaeology, linguistics and genetics* (eds Blench, R., Sagart, L. & Sanchez-Mazas, A.) (Routledge Curzon, London, 2005).
7. Kryukov, M. V. & Kh., S.-I. *Ancient Chinese* (Vostochnaya kniga, Moscow, 2020).
8. Egorova, M. A., Egorov, A. A. & Solovieva, T. V. Modeling the distribution and modification of writing in proto-Chinese language communities. *ADML* **54,** 92–104 (2020).
9. Egorova, M. A., Egorov, A. A. & Solovieva, T. M. Features of archaic writing of ancient Chinese in comparison with modern: historical context. *Voprosy Istorii,* 189–207 (2021).
10. Zinder, L. R. *General phonetics* 2nd (Higher school, Moscow, 1979).
11. Lee, W.-S. *An articulatory and acoustical analysis of the syllable-initial sibilants and approximant in Beijing Mandarin* in *Proceedings of the 14th International Congress of Phonetic Sciences* **413416** (San Francisco, 1999), 413–416.
12. Kodzasov, S. V. & Krivnova, O. F. *General phonetics* (RGGU, Moscow, 2001).

13. *Musical encyclopedia* (Soviet Encyclopedia, Moscow, 1978).
14. Shironosov, V. G. *Resonance in physics, chemistry and biology* (Publishing House "Udmurt University", Izhevsk, 2000).
15. Chion, M. *Audio-Vision. Sound on screen* (Columbia University Press, NY, 1994).
16. Egorova, M. A., Egorov, A. A., Orlova, T. G. & Trifonova, E. D. Methods of research of hieroglyphs on the oldest artifacts – introduction to problem: history, archeology, linguistics. *Voprosy Istorii,* 20–39 (2022).
17. Keightley, D. N. *Sources of Shang history: the oracle-bone inscriptions of Bronze Age China* (Berkeley, London, 1985).
18. *Hieroglyph* 家 *"house, family"* https://en.wiktionary.org/wiki/.
19. *Hieroglyph* 立 *"stand"* https://en.wiktionary.org/wiki/.
20. *Hieroglyph* 交 *"exchange, transfer, give"* https://en.wiktionary.org/wiki/.
21. Pinson, M. H., Ingram, W. & Webster, A. Audiovisual quality components. *IEEE Signal processing magazine,* 60–67 (2011).
22. Urazova, S. L., Gromova, E. B., Kuzmenkova, K. E. & Mitkovskaya, Y. P. Audiovisual media in the universities of Russia: Typology and analysis of the content. *RUDN Journal of Studies in Literature and Journalism* **27,** 808–822 (2022).
23. Carlson, R. Models of speech synthesis. *Proc. Natl. Acad. Sci. USA* **92,** 9932–9937 (1995).
24. Arai, T. How physical models of the human vocal tract contribute to the field of speech communication. *Acoust. Sci. & Tech.* **41,** 90–93 (2020).
25. Story, B. H. & Bunton, K. A model of speech production based on the acoustic relativity of the vocal tract. *J. Acoust. Soc. Am.* **146,** 2522–2528 (2019).
26. Teixeira, A. J. S., Martinez, R. & Silva, L. N. Simulation of human speech production applied to the study and synthesis of European Portuguese. *EURASIP Journal on Applied Signal Processing* **9,** 1435–1448 (2005).
27. Kinahan, S. P., Liss, J. M. & Berisha, V. TorchDIVA: An extensible computational model of speech production built on an opensource machine learning library. *PLOS ONE.* doi:10.1371/journal.pone.0281306 (2023).
28. Maurerlehner, P., Schoder, S. & Freidhager, C. Efficient numerical simulation of the human voice. *Elektrotechnik & Informationstechnik* **138/3,** 219–228 (2021).

Information about the authors

Egorova, Maia A.—Candidate of Political Sciences, Associate Professor at the Department of Foreign Languages of the Faculty of Humanities and Social Sciences of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University) (e-mail: Mey1@list.ru, ORCID: 0000-0003-2931-8330)

Egorov, Alexander A.—Doctor of Physical and Mathematical Sciences, Consulting Professor of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University) (e-mail: alexandr_egorov@mail.ru, ORCID: 0000-0002-1999-3810)

УДК 004.9+086.7+534+621.397+681.84/.85+791.43+811.58
PACS 01.50.F-, 02.60.-x, 29.85.-c, 42.30.Va, 43.38.Md, 43.58.+z, 43.58.Kr, 43.60.+d, 95.75.Pq
DOI: 10.22363/2658-4670-2025-33-3-309-326 EDN: HEVOIT

Исследование иероглифов с помощью методов аудиовизуального цифрового анализа

М. А. Егорова, А. А. Егоров

Российский университет дружбы народов, ул. Миклухо-Маклая, д.6, Москва, 117198, Российская Федерация

Аннотация. Проведённое исследование античных письменных текстов и знаков показало, что иероглифы и строение архаического предложения имеют много общего с современным китайским языком. В контексте истории и эволюции китайского языка подчёркнуты его характерные тональность и мелодичность. Основное внимание в работе уделено исследованию звуковых свойств иероглифов (ключей), встречающихся одновременно в древнейших надписях, а также в современных текстовых сообщениях. В статье использованы современные цифровые методы анализа звуков с одновременной их визуализацией. Для характеристики звучания иероглифов (в соответствии с принятой в Китае фонетической транскрипцией Пиньинь) использованы две (FI, FII), три (FI, FII, FIII) или четыре форманты (FS, FI, FII, FIII), которые создают характерную F-картину. Предложенная нами модель четырёх формант для типовых иероглифов (ключей) названа базовой «F-моделью», она является новой и оригинальной. Для визуализации формант применены программы цифровой обработки звуковых сигналов. Полученные данные сравнивались с соответствующими спектрограммами для мандаринского (стандартного) диалекта китайского языка. Установлено их соответствие друг другу. При анализе F-картин использовалась оригинальная модель, которая позволила охарактеризовать спектрограммы в частотной и временной областях. Дано формализованное описание основных компонентов базовой «F-модели» произношения иероглифов. В заключение отмечено несколько областей, в которых перспективно использование различных методов аудиовизуального исследования: передовые инновационные технологии (искусственный интеллект и виртуальная реальность); телевидение, театральное видеопроизводство; определение качества аудиовизуального контента; образовательный процесс. Проведённое исследование показало, что описанные перспективные методы исследования могут быть полезны при анализе подобных античных иероглифов.

Ключевые слова: анализ речи, китайские иероглифы, спектрограмма, форманты, субгармоники, базовая «F-модель», лингвистика, обработка данных, компьютерное моделирование