Sonderdruck aus

Dmitri Zakharine / Nils Meise (eds.)

Electrified Voices

Medial, Socio-Historical and Cultural Aspects of Voice Transfer

With numerous figures

V&R unipress

ISBN 978-3-8471-0024-9
ISBN 978-3-8470-0024-2 (E-Book)

Contents

E-Voice Synthesis and E-Voice Montage. Techniques and Social Meaning

Andrey Smirnov

Synthesized Voices of the Revolutionary Utopia: Early Attempts to Synthesize Speaking and Singing Voice in Post-Revolutionary Russia (1920s)

Abstract

In a number of articles from the 1930s, composer, musical theorist and journalist Arseny Avraamov, involved in graphical sound research and related methods of sound synthesis, proposed to vocalize the writings of Lenin by synthesizing the author's voice using new technological means. Moreover, in 1943 he argued against the new Soviet anthem, contending that the real revolutionary anthem should be based on new approaches to harmony and performed by the voice of Vladimir Mayakovsky, which was to be synthesized in Avraamov's "Poetical Laboratory". Were these ideas little more than science fiction at the time, or were Avraamov's projects and proposals more immediately viable and was he actively seeking to develop the technology necessary to deliver them? By analyzing forgotten inventions from the 1920s as well as the manuscripts of Boris Yankovsky, the researcher of graphic sound who was a student and follower of Avraamov, the author intends to prove that Avraamov was fully justified to offer the above-mentioned projects. The paper will present examples and reconstructions based on Yankovsky's technique of synthesis and hybridization of sound.

In 1932 composer, musical theorist and journalist Arseny Avraamov proposed to vocalize the writings of Lenin by synthesizing the author's voice using new technological means. He wrote:

> "[…] At all-Union sound conference in 1930 I have mentioned an idea of a conceivable reconstruction of Ilyich [Lenin] voice (according to remained records with the subsequent timbre correction by memory) and, hence, possible vocalization of mute pieces of the Lenin's chronicle, by precise assignment of fragments of the shorthand report uttered by him in each particular moment of speech, and, at last, in general a reconstruction of his speeches which were not recorded in any way, except the shorthand report. […] Our baby just learns to utter 'papa and mommy': we develop graphs of different elements of speech, solving problems which are far outside the boundaries of syntonfilm [graphical sound]. […] But soon we will come solve the problem of 'expressive intonations' which was studied only by 'hack craftsmen', thus there is nothing to utilize. And again—a rhythmic and emotional 'dynamics'—in a new way […]" (Avraamov 1932: 48 – 51).

Somewhat later in 1943 Avraamov argued against the new Soviet anthem, contending that the real revolutionary anthem should be based on new approaches to harmony and performed by the synthesized voice of Vladimir Mayakovsky: "Still we will write the real anthem of the Victory not with helplessly-nurseries Mikhalkov's rhymes, but with an ingenious text (both by the form and content) of new super-Mayakovsky which will come out of my poetic studio in few months" (Rumiantsev 2007: 35). Reflecting on his plans oriented to the synthesis of speech, Arseny Avraamov wrote:

> "[…] the problem of synthesis of speech, […] can be put in three ways:
> 1. 'Portrait'—a reconstruction of individual speech of the given subject.
> 2. 'Typical', impersonal reconstruction of an image of speech—the given nation, a class, a gender, age, character.
> 3. Artistic, synthetic construction of any 'specific' conditional, unknown speech character. […]
>
> Perhaps the easiest solution is to create a certain 'wooden', an 'automatic' voice which will find at once its application in 'animation' [cartoons], to voice any rural hovel or a certain venerable professional bureaucrat at congress. […] The whole series of such grotesque voices will be inevitably created before we will learn how to produce a natural human voice. At this second stage the differentiation of typical features will begin; here the techniques will move far ahead, but practically the second stage gives us not much from the point of view of the specifics of syntonfilm [graphical sound] since we stay at the 'naturalistic' level which does not pay back our creative expenses, and from this level it is not easy to to follow through on the third stage: creations of the artistically comprehended types of speech. There are so many of them that it is not necessary to list; Anyway this idea does not seem feasible by 'natural filming'—beauty, power, expressiveness, comedy, and again a series of grotesques, but already found consciously, with detailed study of the conceived character.
>
> And at last, 'portrait', as the highest stage, which is most difficult and very limited in possibilities. And meanwhile all of us are still in continuous crescendo molto, and climax is still far ahead"(Avraamov 1932: 48–51).

Were these ideas little more than science fiction at the time, or were Avraamov's projects and proposals more immediately viable and was he actively seeking to develop the technology necessary to deliver them?

Speaking Machines

The mid 1930s was a crucial point because of the clash with the two great cultures—the anarchistic artistic and scientific Utopia of the epoch of Projectionism (Pchelkina 2012) of 1920s and the totalitarian, highly centralized Anti-Utopia of 1930–50s. While the official history of Russian speech technologies, based on archives of various famous and State supported scientific schools and counting the beginning in mid 1930s, is fairly well documented, the inventions and discoveries of the community of researchers—representatives of the artistic Utopia of the 1920-s have been largely forgotten.

Figure 1: Ornamental soundtracks, drawn by Boris Yankovsky in 1931 with his own profile (left) as well as the profile of Arseny Avraamov (right). Andrey Smirnov archive.

One of the main features of this generation was cross-disciplinary knowledge and activities. For instance, the objective of the most radical revolutionaries in the realm of music and acoustics was to unite efforts to produce revolution in music theory and techniques based on the cross-connection of arts and sciences (Smirnov 2010: 97 – 99). They declared that academic views of music theory were dull and scholastic, and that techniques related to it were old-fashioned (Sholpo 1939: 248.). As the pioneer of sound synthesis, inventor Evgeny Sholpo put it:

> "Physicists who would begin research of the laws of the generation and distribution of sounds are needed: physiologists are needed who would investigate processes of perception of sounds and their general influence on human body; besides that, both (mainly physiologists and psychologists) should be trained in mathematics to make the results of their research understandable so that the data obtained can be considered as certain factors instead of shapeless, sketchy hypotheses"(Sholpo 1917: 7).

There were two main intentions focused at the development of new sound machines: to play and compose with any sounds at will and to synthesize speech and singing. Quite often both were combined in a single device or method. Because of the lack of information and knowledge some proposals were nothing more, then a repetition of old and well known ideas. For instance in 1929 Russian inventor V.s. Toropov patented the "Apparatus for the reproduction of speech and singing" (Toropov 1929) based on exactly the same principle as the famous Amazing Talking Machine, where, similar to human organs of speech, the air stream crosses a passageway that resembles human vocal cords, and ponder. It was firstly exhibited in 1845 by Joseph Faber at the Musical Fund Hall in Philadelphia.

Nevertheless, many inventors patented new sound machines, based on electro-optical, electro-mechanical, newest electronic, and even audio-computing techniques and were ahead of their time by decades, although in many cases we still have no information about the built prototypes.

Perhaps the most advanced proto-speech synthesizer was the "Mechanical keyboard instrument for the reproduction of speech, singing and various sounds" (Tambovtsev 1925) (fig. 2), patented in 1925 by the inventor and mathematician D. G. Tambovtsev. The instrument was primarily intended for reproduction of artificial speech and singing: each key of its keyboard corresponded to the number of loops of steel tapes equal to the number of vowels of Russian language, and the same number of electromagnets (magnetic heads), and each of these tapes was prerecorded with one particular vowel having a pitch, corresponding to the number of the key on the keyboard, whereas other tapes of the system had no keys corresponding to them, were prerecorded with consonants or pitch-less noises. It was a kind of proto-sampler, very similar to the Mellotron, popular in 1970s.

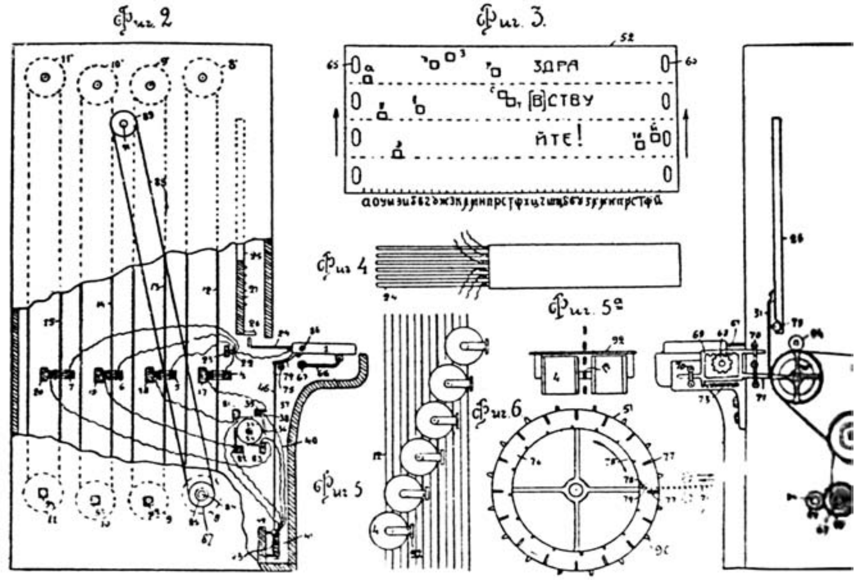


Figure 2: The diagram of the mechanical keyboard instrument for the reproduction of speech, singing and various sounds (Tambovtsev 1925). Andrey Smirnov archive.

To facilitate a kind of *Concatenative synthesis*[1] of speech and singing as well as any other complex sounds the instrument was equipped with a special "sequencer"—the program mechanism, based on a punched tape, capable to play preprogrammed sequences of sound segments or speech phonemes each time the key is pressed, forming words, phrases. The special pressure sensitive musical keyboard was intended to control pitch and volume of sound, permitting to imitate singing.

In fact the method of speech synthesis, based on numerous prerecorded sounds, was not so well known, but still it was quite popular. According to NIMI[2] correspondence, it was offered many times by different inventors during the 1930–40s. Perhaps it was the easiest solution to create a certain 'wooden', an 'automatic' voice (according to Avraamov). Unfortunately this technology did not allow to reconstruct an individual speech by any g subject nor to imitate 'expressive intonations': None of the known proposals could guarantee control over intonation.

1 Concatenative synthesis is based on stringing together segments of recorded speech.
2 NIMI: the Scientific Research Musical Institute at Moscow State Conservatory. The successor of GIMN (the State Institute for Musical Science).

Synthetic Acoustics

In the early 1930s Arseny Avraamov was involved in graphical (drawn) sound research—a consequence of the newly invented sound-on-film technology (Smirnov 2010: 100 – 103), which made it possible to access the soundtrack as a visible graphical trace in a form that could be studied and manipulated. It also opened up the way for a systematic analysis of the optically recorded wave-formswhich could be used to synthesize any sound at will. Avraamov wrote:

> "Since we are real hosts of the soundtrack in each portion of the frame, we simply consider that our method gives us enough means to turn the sound development and dynamics in any desired direction any moment, since we are not limited by the tools of past material culture—musical instruments,—banding our hands by their technical features and limitations" (Avraamov 1932: 48 – 51).

Meanwhile in the autumn of 1930 Arseny Avraamov founded the Multzvuk Group to conduct a research and developments related to the graphical sound techniques. Painter and amateur acoustician Boris Yankovsky belonged to the stuff of the Group either. Yankovskywas responsible for acoustical experimental studies, developing methods for the synthesis of sounds with glissando, timbre cross-fades and polyphony by means of multiple shooting on the same optical soundtrack (a type of multi-track recording).

A year later in the autumn of 1931 the Multzvuk Group moved to NIKFI (The Scientific Research Institute for Cinema and Photography) and was renamed the Syntonfilm Laboratory. In December 1932 NIKFI stopped supporting Syntonfilm and the group moved to Mezhrabpomfilm[3] where in 1934 it was finally closed as it was unable to justify itself economically.

In 1932 Boris Yankovsky has separated from Avraamov's Multzvuk Group. He established his own laboratory in Moscow and created essentially new techniques of spectral transformations and synthesis of various sounds, including a singing voice and speech, confirming Avraamov's foresights on future possibilities of speech and vocal synthesis. The result was an independent control over the pitch and duration of sound on a spectral level, resembling the popular computer music techniques of cross synthesis and the phase vocoder. Yankovsky wrote:

> "[…] I found the idea of synthesis while I was laboriously working on 'drawn sound'. And this is the chain of my considerations:
> The colour of the sound depends on the shape of the sound wave;
> Graphical representation of the sound wave could be analyzed and represented as the Fourier series of periodic functions (sine waves);

3 One of the leading film production companies, created in 1928 in Moscow. In 1948 it was renamed "Kinostudia imeni Gorkogo": Gorki Film Studio.

> Consequently, the sound wave could be re-synthesized back with the same set of sine waves. Nobody did this before the invention of graphical (drawn) sound just because there were no technical means and methodology for sound reproduction from such graphical representations of sound. As with electrons (the neutrons and protons) the number of which defines the quality of the atom, so do sine waves define the quality of the sound—its timbre.
> The conclusion: why not initiate a new science – synthetic acoustics?
> It would make sense if we could define (at least in draft) a sort of Periodic Table of Sound Elements, like Mendeleev's Periodic Table of Chemical Elements. The system of orchestral tone colours has gaps between the rows that could be filled by means of syntheses, like the gaps in the rows of Mendeleev's Periodic Table of Elements have been filled with the latest developments in chemistry. […] It is obvious that the method of selection and crossing of sounds and instruments, which is similar to the method of Michurin [Ivan Michurin, the famous Russian biologist-horticulturist], will give us unprecedented, novel 'fruit-hybrids' that are technically unattainable for a usual orchestra […]" (Yankovsky 1932 – 1940: 15, 45).

Audio Computing and the Rules-based Synthesis in the 1930s

The method developed by Yankovsky was based on pure audio computing techniques and possessed very common properties such as discretization and quantization of spectral data, manipulation with ready-made parts, and operations with selections from databases of the basic primitives (templates) that distinguish it from analogue methods. It can be considered as a sort of proto-computer for music techniques with many of the typical features of modern digital technology of sound and music computing.

Right from the start Yankovsky intended to work with a modified animation stand called the *Vibroexponator*, shooting still images of artificial drawn sound waves by means of a rostrum camera. This meant that the discretization of time scale was predetermined by 24 frames per second, with each successive frame containing one stable sample—a sort of momentary photograph of the constantly changing sound. The audio waveform of each successive frame was calculated according to the spectral content, based on the specific set of spectral templates he had devised.

To perform complex mathematical calculations of waveforms as well as other important parameters of sound and automated musical performance such as rhythm and dynamics, there were special 'employee-computers' on the staff in the laboratories of Boris Yankovsky and Evgeny Sholpo. These were mathematicians whose specific task was to make calculations (fig. 3).

Yankovsky proposed the method, based on research on structural similarities and distinctions in spectra of sounds of different character to limit as far as

Figure 3: ‘Employee-computers’ on the staff in the laboratories of Boris Yankovsky and Evgeny Sholpo. Leningrad, the late 1930s. Courtesy of Marina Sholpo.

possible the number of calculations needed for synthesis of various complex sounds by means of additive synthesis. In order to achieve this he decided to:

analyze the spectrums of various sounds
divide all sounds into classes according to common features of timbres, related to spectrums and spectral dynamics
analyze these common features (formants, for example)
divide the spectrum into groups of overtones responsible for the specific character of the timbre
calculate and draw the waveforms, related to these spectral groups
build a library of drawn waveforms for further manipulations in the framework of various synthesis tasks.

Yankovsky named these final drawn waveforms ‘*spectro-standards*’ or ‘*spectral templates*’, semiotic entities that could be combined to produce sound hybrids, based on a type of spectral mutation. According to Yankovsky, for the final stage of sound synthesis there is no need to deal with spectrums but rather to manipulate the predefined waveforms—the readymade templates—selecting them from the library. These templates could be modified by changing their sizes by optical means (or mathematically and graphically) according to calculated desirable amplitudes and frequencies, mixing the results optically or by means of mathematical calculation of the final waveform.

As a result of his long-term research Yankovsky ascertained that all natural sounds could be divided into classes corresponding to the characteristic types of spectrum as well as to the acoustic properties of sound sources, like:

specific features of the spectral content of a sound (e.g. predominance of even or odd overtones etc.)
presence of formants, reflecting the resonant properties of the sound source
features of transients in a spectrum, especially during the sound attack
specific amplitude envelopes (attacks and decays)
specific amplitude and spectral modulations (e.g. various forms of vibrato)

He also found out that it was possible to allocate properties that were common for most sounds within the boundaries of one class. Yankovsky considered the common formant the most important common feature of different sounds within one class.

Experimenting with optically recorded sounds, Yankovsky noted that transposing them up and down by means of varying the playback speed destroys the character of the sound. According to Yankovsky's analysis:

> "The formant of sound, if it is possible to speak about the formant in this particular case, and most precisely the whole fixed structure of overtones, moves in parallel to the fundamental frequency along the whole range of transposition" (Yankovsky 1932–1940: 71.1).

To solve the problem of pitch transposition keeping the formants fixed to preserve the character of sound, Yankovsky suggested calculating three templates per octave for the whole pitch range of each particular sound, transposing each template by means of varying the playback speed in the range not exceeding one third of an octave. He recollected:

> "Once, synthesizing spectral templates based on the formant of the vowel 'a', I found their similarity to the waveforms of optically recorded woodwinds in their high registers. In other words, the formants in their high registers differed, apparently, only by one frequency zone, not exceeding an interval of a small third. […] To keep the sound character constant it is necessary to synthesize a series of templates possessing the common formant for different pitches, so that the loudest overtones would be located in the same frequency zone […] The question is of the quantity of notes in a range of the instrument for which it is necessary to synthesize spectral templates […] I planned to synthesize 3 templates per octave of each timbre. Certainly, it was a compromise in relation to the requirement of absolute stability of the formant […] but in practice this shift was absolutely unnoticeable" (Yankovsky 1932–1940: 71.2).

According to Yankovsky's classification he was working with four basic levels of spectral organization, related to his specific method:

Simple (or pure) tone—harmonic vibration, based on a sine law
Complex tone—the superposition of several simple tones

Syntone (synthetic tone)—artificial complex tone, based on the mathematical addition of simple tones related to some particular spectral template and created by means of Graphical Sound techniques.

Synthetic Instruments—the library of Syntones, related to some particular sound character (instrument) with fixed formants, intended for music creation, based on Graphical Sound techniques. It has nothing in common with ordinary 'musical instruments' played by hands and fingers (Yankovsky 1932 – 1940: 68).

The Vibroexponator

To produce the sound, dynamically changing in time, one would have to calculate the sequence of static frames, in which each frame represents the successive state of changing timbre. In order to produce a final soundtrack one would have to cross-fade successive overlapping frames by optical means to achieve smooth transitions and to avoid clicks. Yankovsky developed his *Vibroexponator* to realize this process in a single tool.

Although Yankovsky referred to the construction of the Vibroexponator in several articles he never described it in details. Nor have any detailed drawings of its construction yet been found. Nevertheless, according to existing descriptions, the Vibroexponator had several stages that enabled the whole process of *syntone* production as well as the final soundtrack, related to one frame.

In January 1939 Nikolai Garbuzov, head of the NIMI Institute at Moscow State Conservatory, sent the researcher Nikolai Zimin to Yankovsky's laboratory to check the state of developments. In his notes 'About the Laboratory for Synthetic Sound' Zimin (1939) wrote:

> "The Vibroexponator is a complex, bulky tool for the optical recording of synthetic sounds to the sound track of ordinary 35 mm film by means of specially produced intensive negatives. The instrument is partly mechanized and provides various motions to the original negative. The automation of the direction control is partially broken and requires extra repairs and maintenance. [...] The slide copying tool is intended for production of intensive negatives from films with transversal soundtracks. It too is a massive construction. The gearbox has at least a 100-fold safety factor and a greater power".

A crucial part of the Vibroexponator was the *slide copying machine tool* (fig. 4) devised to convert the initial 'transversal' optical soundtrack (produced by pure drawn methods based on mathematical calculations of the spectral templates) into the 'intensive' form necessary for further processing. To produce the *slide copying* a film was placed in the mount of the copying cassette with a thin aperture located along the film enabling light to pass through onto the photographic plate behind. An extended variable density image of the original transversal waveform was produced by dragging the cassette down in front of the unexposed photographic plate (fig. 5 – 6).

The 'intensive' (variable density) image of the sound waveform could then be used to produce new waveforms related to different lower pitches of the sound. To do this, the photo plate with variable density waveform needs to be mounted behind another thin aperture, rotated according to the scale with precisely calculated angles of rotation related to desirable pitch shift (fig. 7). Further film is then exposed using this aperture. The rotation of the original has the effect of stretching the waveform without changing any of its relative characteristics (fig. 8). This part of the Vibroexponator was called the 'Syntone Exponator'.

The Vibroexponator was mounted as an extension of a rostrum camera. In this way, successive frames of film could be exposed to build spectral transformations by incremental changes in pre-synthesized waveforms—each increment being one frame of film. The change in sound was enabled by cross-fades between successive frames. Each frame was faded into the next to make edits in the sound less audible. This was realized in two ways, the first of which involved defocusing towards the edges of each frame, creating a blurring, or smoothing of the waveform. The second involved processing each frame with a bell-like amplitude envelope, based on one period of the sine wave (fig. 9 – 10), by means of a special mask located in the top part of the Vibroexponator.

Another part of the Vibroexponator incorporated a special multi-segment mask to produce fast envelopes with discretization, equal to 3 steps per frame, to produce amplitude and spectral vibrato (fig. 11 – 12). The final processing included using the top part of the Vibroexponator to produce slow envelopes for cross-fades between two different *syntones.*

To manipulate the dimensions of the image of the initial 'transversal' soundtrack, related to the pitch and amplitude (for example, to double the pitch), different optical methods were utilized, including the *Anamorphot* optical tool, based on an anamorphic lens system that was developed at the Leningrad Institute for Precise Mechanics and Optics in response to Yankovsky's idea.

Syntone Database

In 1933 Yankovsky was invited to the Mosfilm Studios to organize the Laboratory for Synthetic Sound Recording where in 1934 – 35 he recorded a sizeable collection of samples of instruments from the Symphony Orchestra of the Bolshoy Theatre as well as sounds of speech, including collections of vowels and consonants. By 1936 the collection of 110 synthesized templates—*syntones* had been created. In 1938, during his experiments with violins at the First Factory of

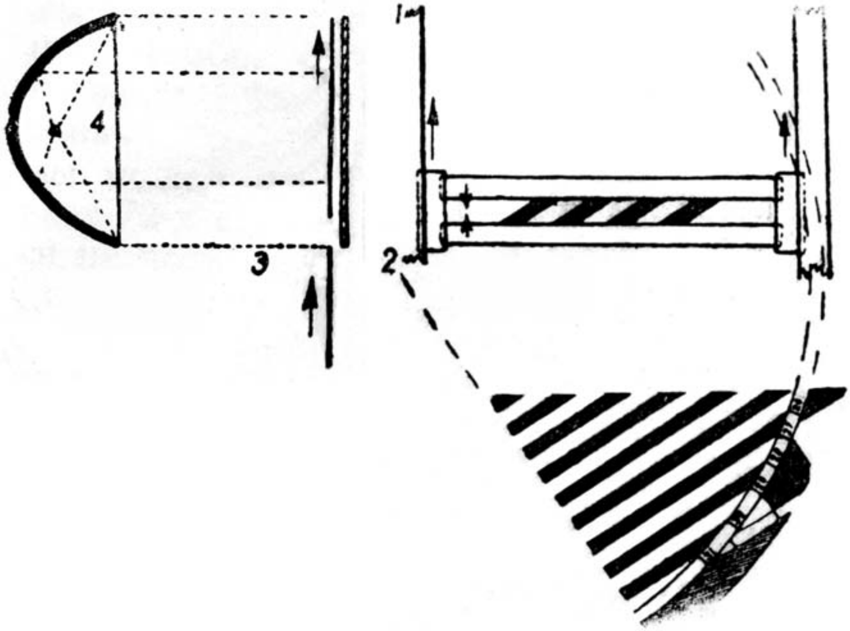


Figure 4: The diagram of the *slide copying machine tool of the* Vibroexponator. Illustration by Boris Yankovsky, 1933. Andrey Smirnov archive.

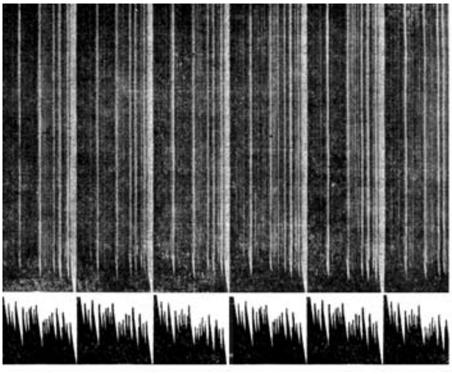


Figure 5: Result of transformation of the 'transversal' soundtrack into the 'intensive' (variable density) one. Illustration by Boris Yankovsky, 1933. Andrey Smirnov archive.

Bow Instruments in Moscow, Yankovsky had the opportunity to work with the new western spectrometer (the first one imported to the USSR) to check his theoretical conclusions. According to his estimations, he had produced and collected hundreds of spectrums. After inspection of Yankovsky's laboratory in 1939 the scholar and researcher Nikolai Zimin (1939) wrote:

Figure 6: Variable density images of the waveforms, produced by Boris Yankovsky in the middle 1930s. Andrey Smirnov archive.

> “As auxiliary material there are several dozen films with a length of 0.5 m each, with multiple periods of synthetic sound curves. In this category there are also about several dozen films with defects. The most ingenious negatives of sound curves adapted to work in the Vibroexponator are big square variable density photo plates and are collected in special albums. As a final result there are about one hundred 35 mm films of lengths from 1.5 up to 20 metres. The shorter films are glued together in long infinite loops with a length of about 20 metres. Several of them were played by Yankovsky, including the clarinet, trombone, nose-like, equal-amplitude complete, equal-amplitude unclear[4], transformations between sounds and special timbres with the addition of high formants. The sound quality as such is not the best and can be significantly improved”.

Unfortunately the above mentioned collection of films and photo plates is not discovered yet except one short film with sound experiments, based on ‘pen-

4 Yankovsky’s terminology refers to the common spectral characteristics of some classes of sounds. ‘Clarinet’ means the class of spectrums with dominant odd overtones; ‘trombone’ and ‘nose-like’ are classes of spectrums with some specific formants; ‘equal-amplitude complete’ means spectrums based on complete sets of overtones (including both odd and even harmonics) with equal amplitudes; ‘equal-amplitude unclear’ means spectrums with some missed overtones. The remaining overtones have equal amplitudes.

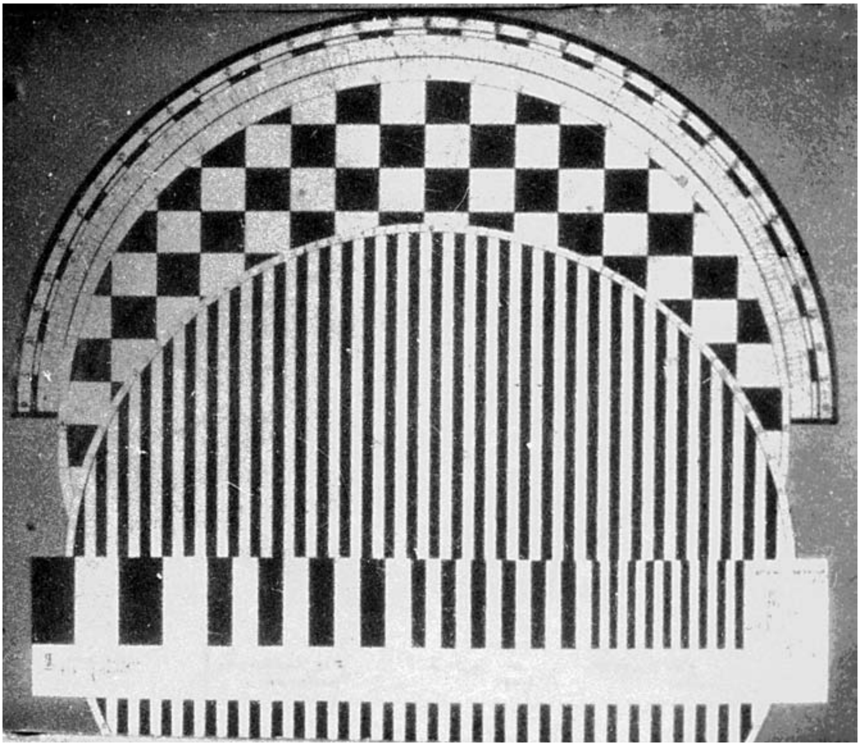


Figure 7: The scale with precisely calculated angles of rotation related to desirable pitch shift. Yankovsky, middle 1930s. Andrey Smirnov archive.

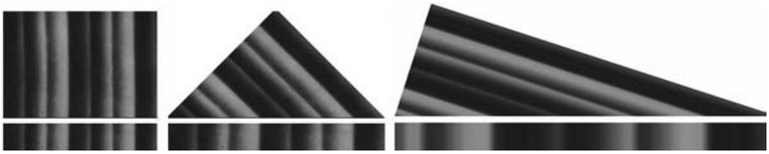
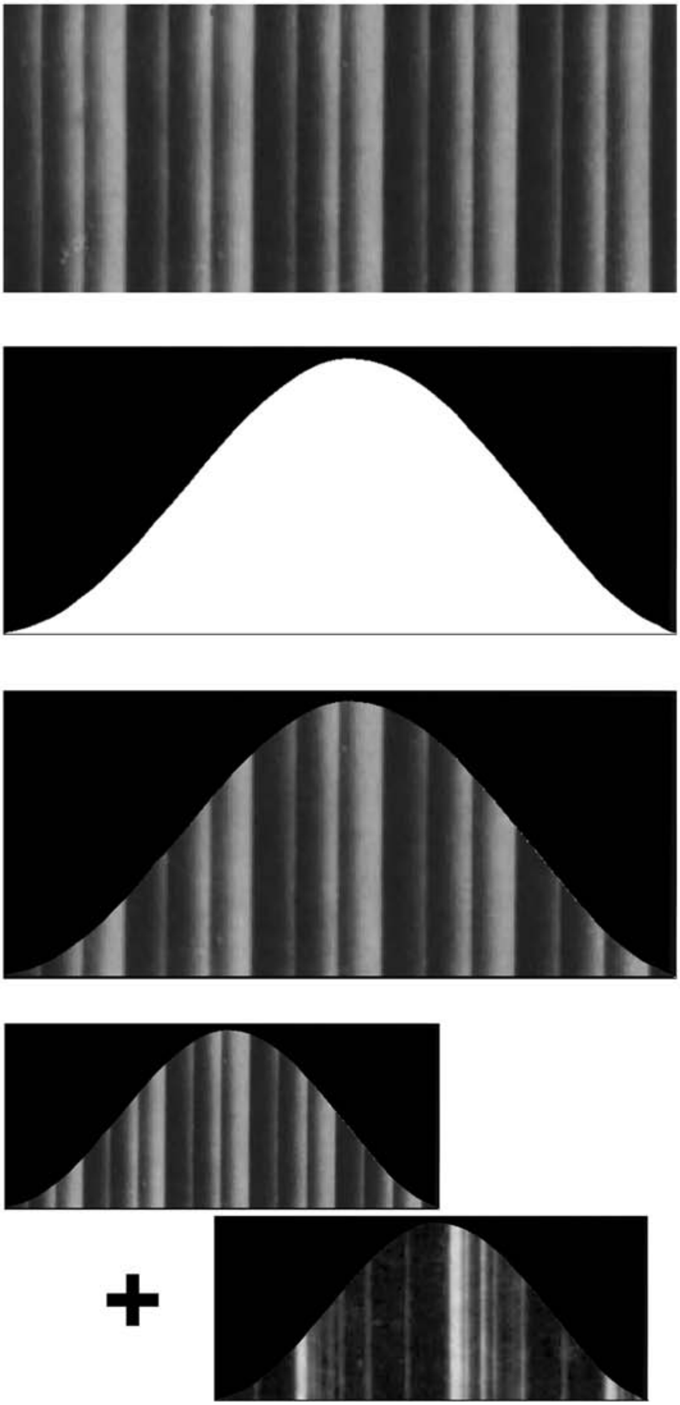


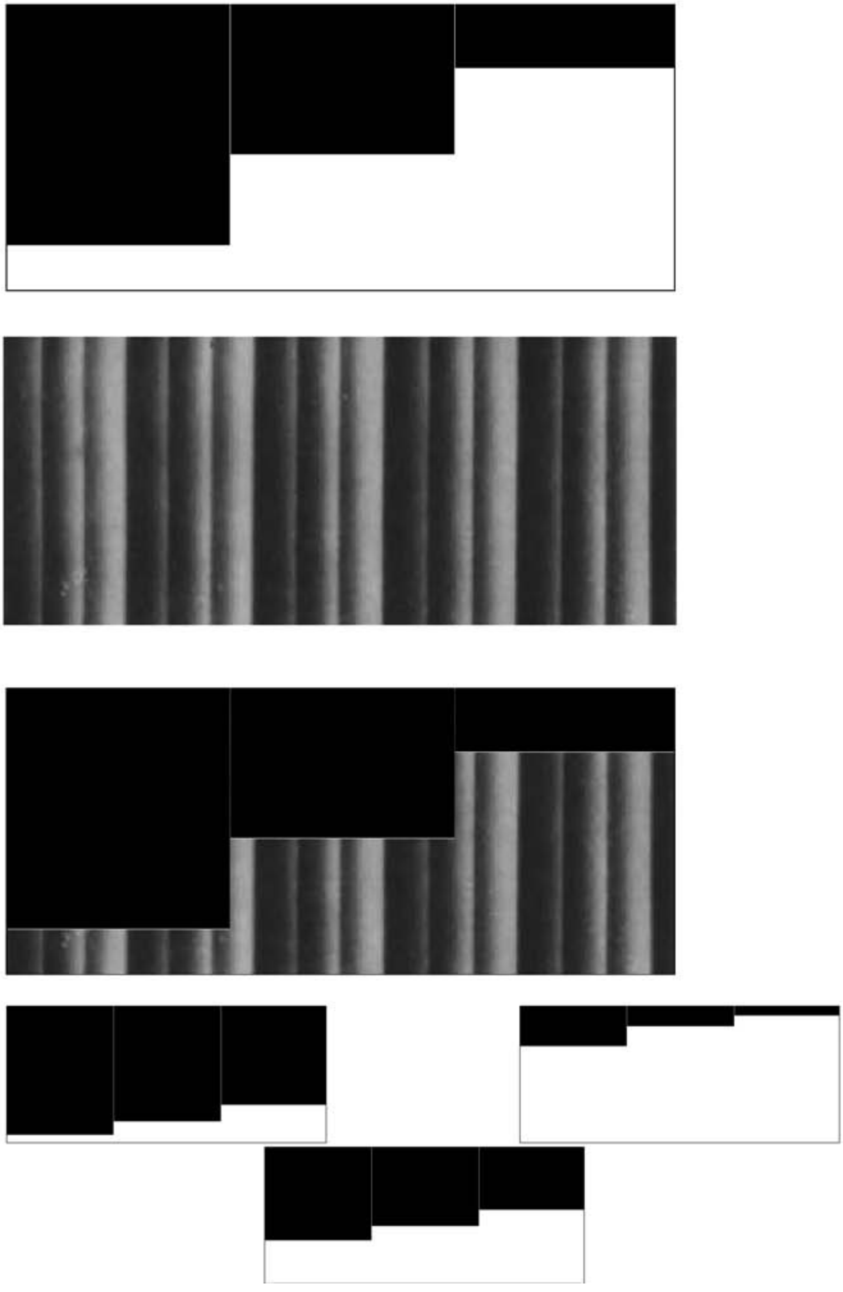
Figure 8: Explanation of the optical time stretching of the variable density soundtrack. The diagram by A. Smirnov.

taovertones'—syntones with sharp formants, very much reminding early computer music experiments with synthesized voices by Newman Guttman at Bell Laboratories in 1957 by means of Music4 software. Nevertheless several dozen illustrations with syntones were included by Yankovsky in his manuscript "Acoustical Syntheses of Musical Colours", which was planned as one chapter for the book "The Theory and Practice of Graphical Sound", written by Evgeny Sholpo in 1936 – 1939. The two were never published. Fortunately all illustrations are available and all syntones are digitized and explored by the author.

One of these illustrations contains a collection of syntones, produced in 1935



Figures 9 – 10: Explanation of the optical crosfading by means of the bell-like mask. The diagram by A. Smirnov.



Figures 11 – 12: The special multi-segment mask to produce fast envelopes with discretization, equal to 3 steps per frame, to produce amplitude and spectral vibrato. The diagram by A. Smirnov.

and based on variations on Russian vowels *А, Э* and *О* (fig. 13). In this collection Yankovsky represented several stages of additive synthesis of syntones with different sets of overtones, starting from a single sine wave as well as different formant structures, based on a single formant and two formants. For instance, on fig. 14.1 and fig. 14.2 are represented drawn single periods of syntones of the phoneme *a* [a] as well as related spectrums. On the spectrogram of syntone Fig.14.1 one can find two formants with frequencies 800 and 1200 Hz. Syntone fig. 14.2 (octave down) has slightly lower formants 700 and 1000 Hz. Both syntones are close to Russian [a], although the second formant is a bit lower, than typical one. Syntones fig. 14.3 and fig. 14.4 illustrate a very successfully synthesis of the vowel э [e]. Though their frequencies differ on an octave, the sharp formants in both cases are equal to 400 and 1600 Hz which is very close to expected values. The syntone *o* [o] (fig. 14.5) is also very successful. It possesses formants 470 and 800 Hz and sounds very natural.

A very special case is collection of the so called 'Pentaovertones' (fig. 15). Yankovsky did an attempt to synthesize narrow band spectrums, reminding single sharp formants. His idea was to calculate waveforms related to sets of five high order overtones, grouped around some particular middle frequency with missed low order overtones, including the first harmonic. His expectation was that stringing together calculated single periods of Pentaovertones, separated by silent segments with predefined length, will permit him to synthesize sounds with different pitches, dependant on the duration of intermediate silent segments, while the frequency of the formant will be fixed and equal to the above mentioned middle frequency. This idea is illustrated on the fig. 14.6 – 8. On the spectrogram one can see the group of four overtones instead of expected five as well as a strong first harmonic, which was expected to be missed. This is a consequence of a distortion, produced by the silent pause between separated periods of syntones. In fact the waveform for the whole period including the intermediate pause should be calculated according to the spectrum desired. Nevertheless it is obvious that even in this case the idea is still valid: we see the strong fixed formant, independent of the pitch, while the spectral content depends on the basic frequency. This idea gives possibility to vary the pitch as well as to construct complex formant structures independent of the pitch with a minimum of calculations.

While the method of synthesis of vowels and various resonating sounds was developed in details, Yankovsky did not offer any approaches to synthesize consonants, hushing sound, voiced consonants and fricatives. In the meanwhile one has discovered a collection of soundtracks related to audio samples of consonants and hushing sounds among the survived photo documents (fig. 16). Perhaps Yankovsky was just at the beginning of his research related to speech and singing voice. Nevertheless, the syntone method as well as the construction

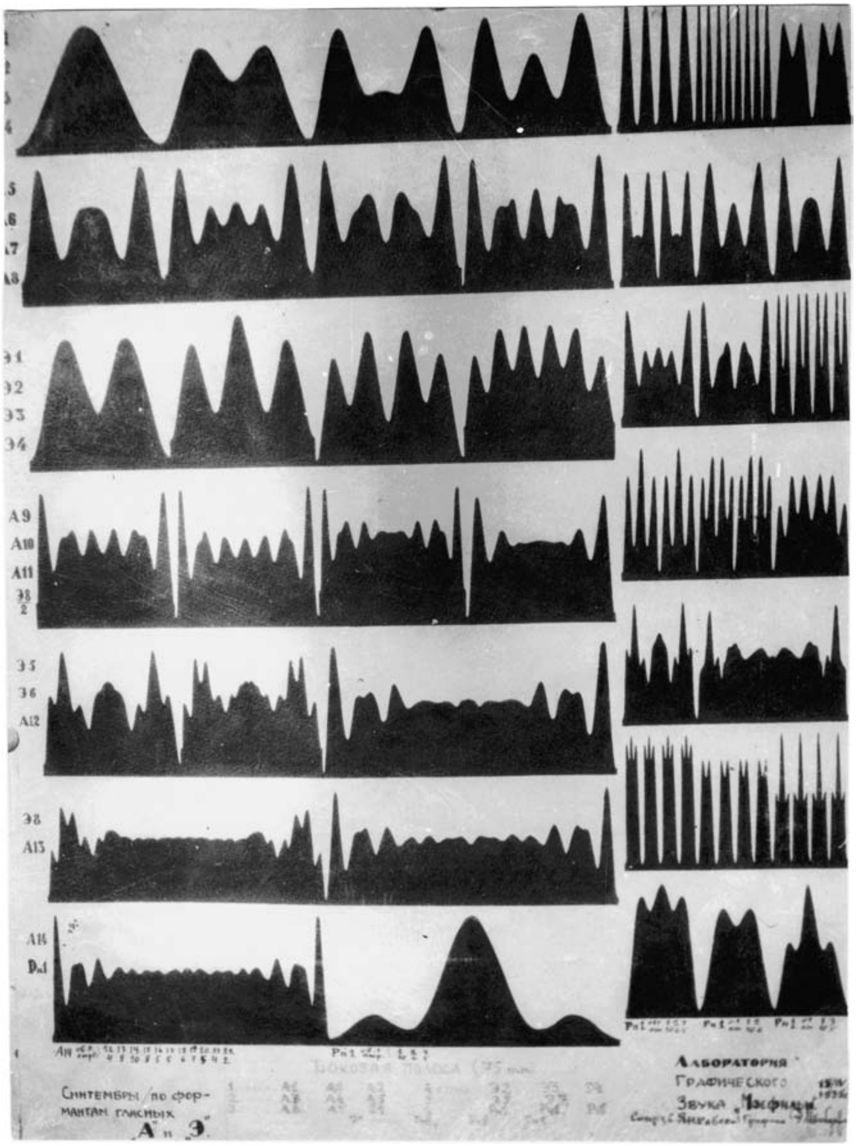


Figure 13: Collection of syntones, produced by Yankovsky in 1935 and based on variations on Russian vowels *А, Э* and *О.* Andrey Smirnov archive.

of the Vibroexponator and related method of audio editing permits to combine both—the rules-based synthesis, requiring audio computing, and the *concatenative synthesis* based on stringing together segments of recorded sounds. Thus in theory, according to the method, developed by Yankovsky, a idea of

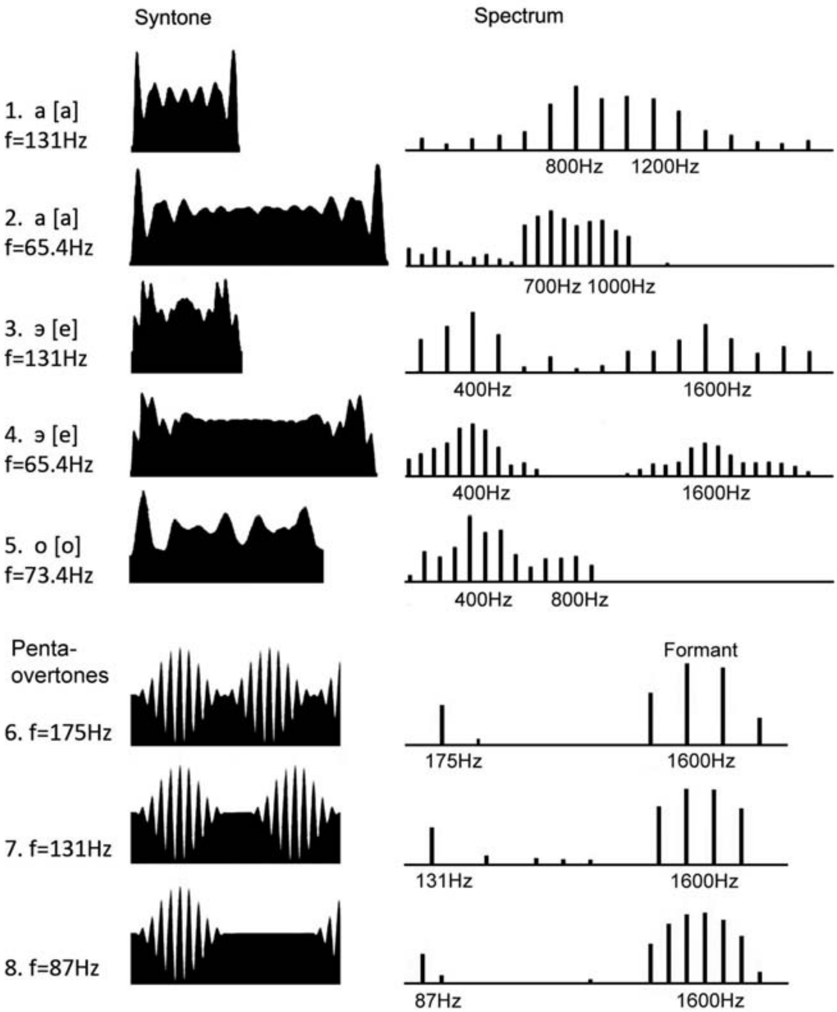


Figure 14: Some syntones and their spectrums. The diagram by A. Smirnov.

reconstructing individual speech or to imitate expressive intonation was almost achievable. Although to get at least very rough results without even worst computer, researcher had to spend enormous efforts and time, incomparable with a single human life.

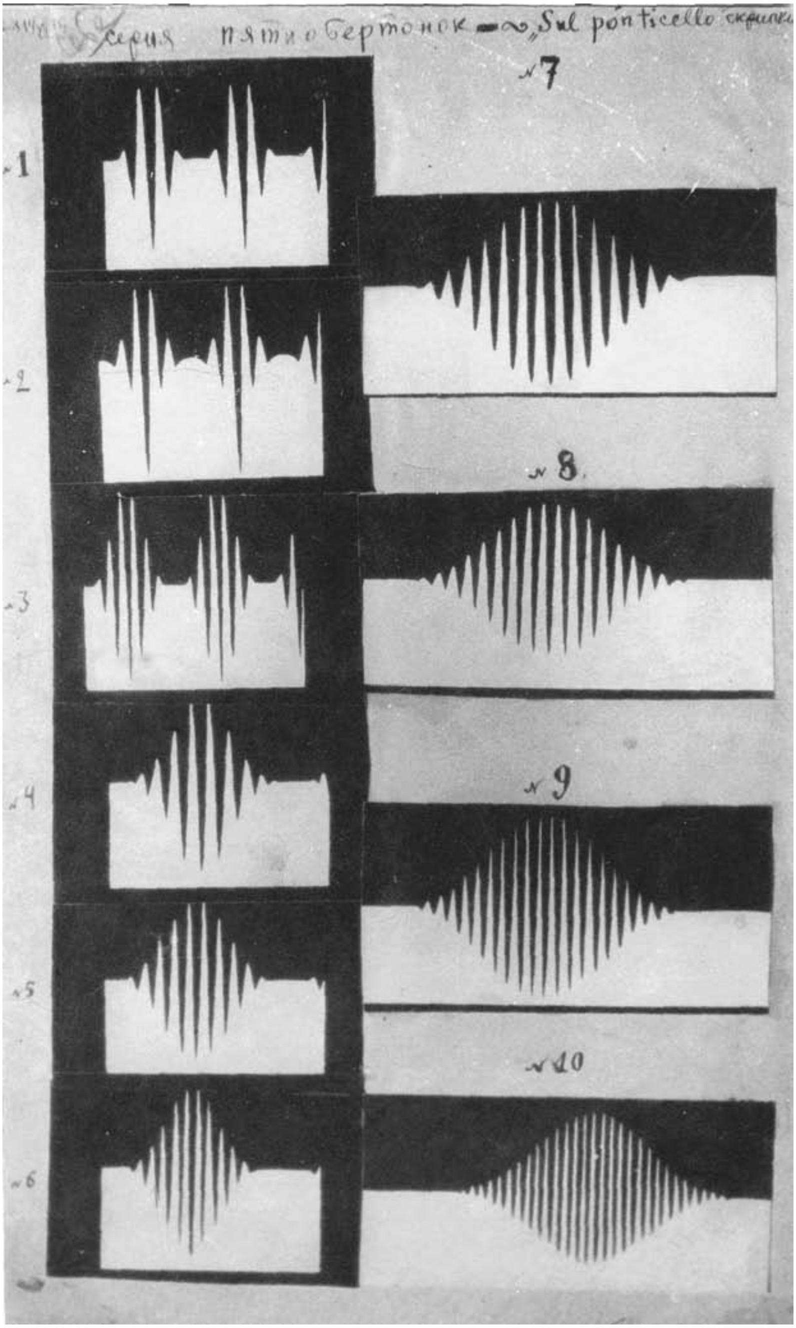


Figure 15: Collection of the 'Pentaovertones'. Yankovsky, 1935. Andrey Smirnov archive.

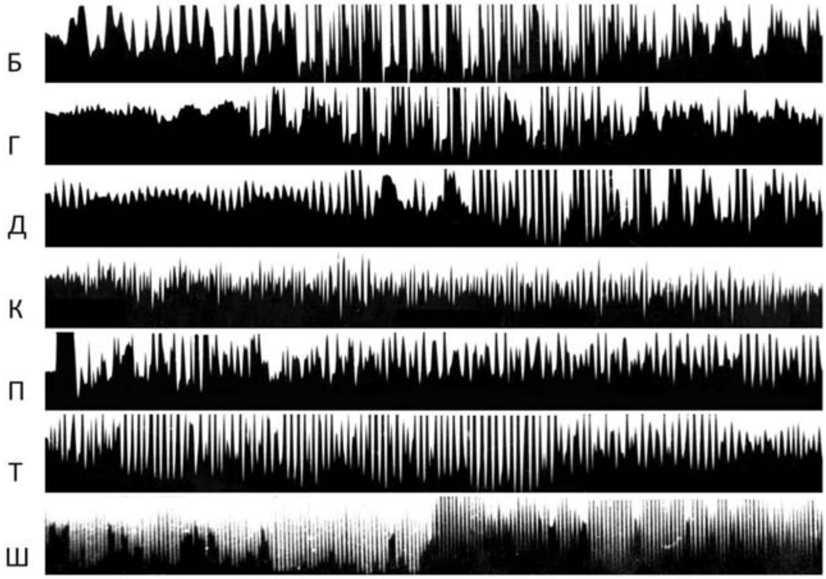


Figure 16: Collection of optical soundtracks with the audio samples of consonants, hushing sounds and voiced consonants. Yankovsky, 1935. Andrey Smirnov archive.

Figure 17: Boris Yankovsky, Leningrad, c. 1939. Courtesy of Marina Sholpo.

Timeline

In 1935 during his work at Mosfilm Studios Yankovsky joined the Autonomous Research Section (ANTES) at the Union of Composers in Moscow, founded by Boris Krasin, Arseny Avraamov and Alexei Ogolevets. After the death of Krasin in 1936 ANTES was closed and the Ministry of Culture stopped funding Yankovsky's laboratory. It was passed to the NIMI institute at Moscow State Conservatory. In 1937 the young NIMI employee, inventor Andrei Volodin[5] developed and built the electronic part (audio amplifiers) of the *Vibroexponator* and Yankovsky finally, after 6 years of research, got his *syntones* sounding.

In January 1939 Yankovsky decided to join the new Laboratory for Graphical Sound, established in Leningrad by Evgeny Sholpo. The main activities of the laboratory were focused on the recording of new *syntone*-based *synthetic instruments*. Yankovsky moved to Leningrad where he was provided with accommodation by the Institute of Theatre and Film. He expected to complete the final version of his *Vibroexponator* in 1940 but his work was curtailed by World War II. During the war Yankovsky and his family were evacuated to Alma-Ata. When Yankovsky found an opportunity to move back to Moscow in 1949, he switched to scientific research work on the acoustics of violins and the history of Graphical Sound was almost forgotten. In the era of analogue synthesis and magnetic tape recording nobody was interested in 'old-fashioned' technology.

Acknowledgments

I would like to thank Lev Bolotsky and Marina Sholpo for unique documents which they rescued from destruction and kept for so many years. I also thank Matthew Price and Rob Mullender for their help.

Andrey Smirnov
Moscow, September 2011

References

Avraamov, A. (1932). Syntonfilm. *Proletarsoye kino:* 9 – 10.
Rumiantsev, S. (2007). *Ars Noviy ili dela I prikluchenia bezustalnogo kazaka Arsenia Avraamova.* Moscow: Deka-BC.

5 Inventor of the Ekvodin synthesizer.

Pchelkina, L. (2012). *Biomechanics of Voice and Movement in Solomon Nikritin's Projection Theatre*. In this volume.

Smirnov, A. (2010). Boris Yankovsky. Leben im Klangspektrum Gezeichneter Klang und Klangsynthese in der Sowjetunion der 30er Jahre. In: *Klangmaschinen zwischen Experiment und Medientechnik*, edited by D. Gethmann. Bielefeld: Transcript, pp. 97–120.

Sholpo, E. (1939). The Artificial Phonogram on Film as a Technical Means of Music. *Collection of works of the Research Institute for Theatre and Music*. Leningrad.

Sholpo, E. (1917–1918) Vrag muziki (The Enemy of Music). *Manuscript. 1917–1918*. Marina Sholpo personal archive.

Toropov V. S. (1929). *Apparat dlia vosproizvedenia rechi I penia* (Apparatus for the Reproduction of Speech and Singing). Copyright certificate N 36345 filed October 26th, 1929.

Tambovtsev D. G. (1925). *Klaviaturniy mehanicheskiy instrument dlia vosproizvedenia zvukov I rechi* (Mechanical Keyboard Instrument for Reproduction of Speech, Singing and Various Sounds). Patent No. 6309 filed May 9th, 1925.

Yankovsky, B. (1932–1940). *Teoriya I praktika graficheskogo zvuka. Akusticheskiy sintez muzikalnih krasok* (The Theory and Practice of Graphical Sound. Acoustical Syntheses of Musical Colours), Leningrad. Unpublished manuscript. Theremin Centre Archive.

Zimin, P. N. (1939) *About the Laboratory for Synthetic Sound*. Notes for the head of the Acoustical Laboratory at Moscow State Conservatory, Professor Garbuzov. Moscow, January 8th, 1939. Theremin Centre archive.