On the Synthesis of Some Artificial Sounds and Words of Human Speech

Viachaslau Vladimirovich Mitsianok

Department of Engineering, Palessie State University, Pinsk, Belarus

Email address:

mitsianok@mail.ru

Abstract: The paper describes the results of numerical experiments on the decomposition of some sounds and words of a person's speech into separate waves with slowly drifting amplitudes, frequencies, phases and their reverse summation in order to identify factors that are both important and not important for automatic speech recognition. The objective of this study is investigation the mathematical features of various sounds and words of human speech without using the method of Fourier transforms. Instead of Fourier transforms, the approximation method developed earlier by the author is used. This method allow expand of periodic or almost periodic functions to sum of modes with slowly varying (drifting) parameters - amplitudes, frequencies, phases. Such decompositions were carried out for samples of vowel sounds, simple syllables and words. After that, the reverse summation of the drifting modes was carried out. Before summation the modes, their parameters were deliberately distorted in order to identify factors, both significant and insignificant for the essence of sounds. The functions obtained in this way are of the nature of artificial sound functions It turned out, that for vowel sounds amplitudes of modes may be averaged over long time without lost the essence of sounds. The phases of sounds may be changed by adding any random constant value without lost their essence too. It has been found that In many cases, for to find the parameters, it is convenient use not the sound function itself, but its time derivative. It was shown, that amplitude of summing modes of sound function may be represent as sum of several Gaussian function as for simple sounds, as for syllables. The appropriate mathematical formulas and tables of parameters of artificial sound functions presented

Keywords: Speech and Person Recognition Throw Voise, Speech Technologies, Data Processind, Fourier Transform, Transform Voice, Expand the Quasiperiodical Sygnals into Base Frequencies

1. Introduction

Despite enormous efforts and financial investments, the problem of speech recognition still does not have a satisfactory solution [1]. The famous American researcher Marvin Minsky recently argued that the pace of progress in speech recognition has recently slowed down, and some of the achievements, which nevertheless took place, were not achieved as a result of new breakthrough ideas, but as a result of an increase in the technical capabilities of computers - speed, memory growth, etc. Perhaps the reason for the slowdown in progress lies in the fact that the approaches and methods that were used earlier have already worked out their resources and it is necessary to do something fundamentally new to move forward further? Perhaps the reason for the slowdown in progress was the use of Fourier transforms? At present some investigators prefers use not Fourier transforms, but other ways [2, 3]. In spite of that, nevertheless majority of investigators use method of Fourier transformations (See [4-11] and the literature, indicated there).

As it well known, the Fourier transformations method, used for the analysis of (quasi) periodic signals has a number of significant drawbacks [4-7, 12-16]. The signal spectra are blurry (in quantum mechanics, this circumstance is the mathematical background of the uncertainty relation), the degree of blur depends on the duration of the signal segment - if the duration is too short, the blur of the lines becomes so great that adjacent lines of the spectrum can absorb each other. If the duration is too long a lot of false lines appear on the spectrum. The longer the signal, the more of these lines. False lines are present in spectrograms even in the case of ideal harmonic signals specified for a limited period of time.

All this also affects the tasks of automatic recognition of human speech and verification and identification of a person by voice. An indirect sign that the Fourier transform method is not suitable for solving these problems is that, despite numerous efforts, serious financial investments, these problems still do not have a satisfactory solution.

In this regard, in [6, 7, 12-16], an approximation method was proposed, which is designed to solve the same problems, but which does not have the inherent drawbacks of the Fourier transformations method. A number of fundamental results were obtained based on the approximation method. It turned out that in the spectrum of individual, long pronounced sounds, there are half-integer (with respect to the baseline) frequencies acting in "bursts", there is a "hard" modulation of the amplitudes of higher modes by the base frequency. Modulation not continious, but broken. Thus, an explanation was found for the failures of the Fourier transform method.

In connection with the certain successes of the approximation method, it makes sense to apply it to create artificial sounds and words of human speech.

$$S = \sum_{i=1}^{n}\left[y(t_i) - y_1(t_i)\right]^2 + \alpha\sum_{k=1}^{n-1}(b_{0,i} - b_{0,i+1})^2 + \alpha\sum_{k=1}^{l}\sum_{i=1}^{n-1}(a_{k,i} - a_{k,ii+1})^2 + \alpha\sum_{k=1}^{l}\sum_{i=1}^{n-1}(b_{k,i} - b_{k,i+1})^2, \quad (1)$$

where $y(t_i)$ —is a time-dependent approximated function describing a signal given by its values at n consecutive times from t_1 to t_n, and

$$y_1(t_i) = b_{0,i} + \sum_{k=1}^{l}\left[a_{k,i}\sin(\omega_k t_i) + b_{k,i}\cos(\omega_k t_i)\right], \quad i=1\ldots n \quad (2)$$

is approximating function, $b_{0,i}$- drifting zero (origin), $a_{k,i}$, $b_{k,i}$ - drifting amplitudes of sine- and cosine- waves (parameters of the approximating function), - ω_i their carrier frequencies, l - number of waves (modes) in the approximating function. In (1) and (2), for simplicity, we can accept $t_i=i$.

Functional (1) is designed as a sum of terms of two types: terms that do not contain the parameter α are responsible for the proximity between the approximated and approximating functions, terms containing α are responsible for smoothing the jumps of the drifting amplitudes of waves of the approximating function when passing along the time axis between adjacent sampling moments.

The larger the value of α is chosen, the smoother the wave amplitudes will be. By calculating partial derivative S (formula 1) with respect to $a_{k,i}$, $b_{k,i}$ and $b_{0,I}$ for all k and i and further equating the results to zero, we obtain a system of linear algebraic equations with respect to the parameters of the approximating function. Having solved this system, we find these parameters. Then this parameters may be substituted in (2) and thus we will expand the approximated function into the sum of waves with slowly varying amplitudes. The resulting approximating function may called reconstructed sound.

If we then subtract the reconstructed sound from the original (approximated) sound and subject the difference to Fourier transforms, it turns out that there are often some

If artificial words and sounds be create, then it will become clear for what exactly should be paid attention to during automatic speech recognition, what features of sound signals make it possible to distinguish one speaker from another, and which ones, on the contrary, have no meaning, they are random, introduced by imperfection of the human speech-making apparatus, they only "get underfoot", distracting the attention of researchers and forcing them to scatter their efforts.

The solution to the problem of generating sounds and words should be started precisely with identifying the mathematical features of various sound units of human speech.

Deciphering the mathematical features of various human speech sounds is also the key to deciphering the mathematical features of speech sounds and other creatures living on Earth, primarily dolphins, elephants, whales, and, in a more distant perspective, to understanding their semantics

2. Approximation Method

The method is based on the functional [6, 7, 12-16].

other carrier frequencies that were not noticed during the first expansion in the Fourier series (integral) due to the small intensity of the modes they carry.

In particular, by this way in [12] it was found that in the spectrum of many sounds there are half-integer (with respect to the base) carrier frequencies. Accordingly, the sound contains whole and half-whole, albeit low-intensity, modes.

Each of the modes included in (2) can be rewritten in a physically more informative form:

$$a_{k,i}\sin(\omega_k t_i) + b_{k,i}\cos(\omega_k t_i) = c_{k,i}\sin(\omega_k t_i + \phi_{k,i}),$$

$$k=1\ldots l, \; i=1\ldots n \quad (3)$$

Then the approximating function looks like this

$$y_1(i) = b_{0,i} + \sum_{k=1}^{l}c_{k,i}\sin(\omega_k i + \phi_{k,i}) \quad \text{I} \quad (4)$$

Here $c_{k,i}$ is the drifting total amplitude of the wave (mode), $\phi_{k,i}$ is the drifting phase. Everywhere below, the term amplitude will be understood as the total amplitude

3. Synthesis of Artificial Mono Sounds

We studied those vowel sounds that could be pronounced for a long time - these are the sounds (mono sounds) "A", "O", "U", "E", "Y", "I", received from several respondents, womens and mens. The sounds were decomposed into modes by a proportional catching network [6] and then restored. In all cases, the reconstructed sound sounded the same as the original sound.

In order to answer the question of what exactly makes the sound "A" as namely sound "A", the sound "O" as namely sound "O", etc., before summation (4), mathematical experiments were carried out for the purpose of deliberate distortion of amplitudes and phases.

First, the phases of all integer modes, except for the basic one, were replaced by artificially calculated ones related to the phase of the basic mode by the formula

$$\phi_{k,i} = k\phi_{1,i} + r_k \ , \ k=1\dots l, \ i=1\dots n. \qquad (5)$$

Here k is the mode number, $\varphi_{1,i}$ is the time-dependent phase of the base mode, r_k is an array of arbitrary numbers. The phase of base mode not changes. Amplitudes of all modes not change too. As it turned out, the sounding of sounds not change.

Secondly, let us pay attention to the behavior of the drifting amplitudes of one of the sound samples "A" (Figure 1).
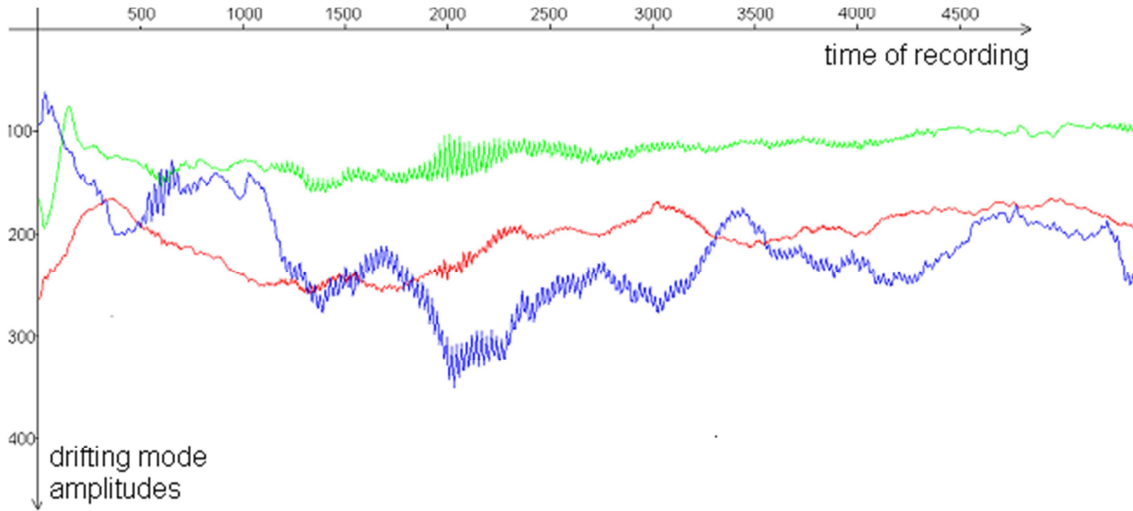
Figure. 1. *Behavior of first 3 lower amplitudes of whole modes of sound "A". The red line is the amplitude of mode No. 1, the green line is the amplitude of mode No. 2, the blue line is the amplitude of mode No. 3. Since the sound was recorded at a sampling rate of 44100, here and everywere below 1/44100 part of a second is taken as a unit oh time.*

As can be seen from Figure 1, the drifting amplitudes vibrate, as it were, chaotically around their mean values. The very fact of the chaotic behavior of general amplitudes suggests that chaos is something introduced that has nothing to do with the individuality of sounds. And so it turned out.

Drifting amplitudes can be averaged over a segment of the sound, and then the actual drifting amplitudes can be replaced with their averaged values, which are constant over the entire segment of the sound. The sound obtained after this amplitude distortion sounded the same as the original.

Third, it turned out that when the modes are summed, the drifting zero and half-integer modes can be omitted. By this rejection, the sound did not change. But if the phase of each of the modes, including the base one, is replaced with a constant but random number throughout the entire sound segment, then the sound quality deteriorated significantly. Instead of a clear sound, what could rather be called the sound of a buzzer was heard.

In search of explanations for this phenomenon, the following mathematical experiments were performed. The averaged amplitude spectrum of each of the studied sounds was combined in formula (4) with drifting phases from any other of the same sounds and from any of the other respondents. After this operation, the sound did not change, sounded clear and corresponded precisely to the amplitudes

So what explains the deterioration in sound quality when replacing drifting phases with constants? It turned out that in all cases the real phases are not strict constants, but drift (float) around some average values with an unstable period from 1.5 to 2.5 Hz and an unstable amplitude of 0.5-2 radians. They are, as it were, "spoiled."

In this regard, the assumption arose that this is how it should be. That the listener's brain is already ready for the fact that the speaker will produce a signal with a corrupted phase, and the sound with an uncorrupted phase is not perceived by the listener's brain as a sound. This assumption came true. When a chaotically changing (within certain limits) value was taken as a phase, the sound again sounded clearly and recognizably

Summing up all of the above, we find that for the synthesis of the above sounds, instead of (4), as one of the options, we can take the formula

$$y_1(i) = \sum_{k=1}^{l} c_k \sin(\omega_k i + kp \sin(i/3300) + r_k), \ i=1\dots n \quad (6)$$

where the averaged values of the amplitudes are given in the following table 1, ω_k are the carrier frequencies proportional to the base frequency, the value of which is given in the last line of table 1, r_k is an array of arbitrary numbers, n is the length of the sound segment (in sampling counts).

The p factor in (6) can take any value in the interval [1, 10], but the best sound is observed when $p = 2$ for the sounds "E", "Y", and $p = 4$ for the sounds "A", "O", "U", "I". The

author's voice was taken as the basis for obtaining the averaged total amplitudes in Table 1. It is also accepted in

(6). The internal sine in (6) provides phase damage. (Other options for phase damage are also possible).

Table 1. *Values of amplitudes of different modes of simple vowel sounds.*

Mode number	A	O	U	E	Y	I
1	637	613	1060	566	1757	914
2	375	714	814	540	354	112
3	674	836	303	1007	65	22
4	794	495	0	61	0	0
5	753	51	0	114	25	0
6	180	0	0	123	51	0
7	49	0	0	90	140	0
8	19	0	0	97	32	0
9	15	0	0	183	54	16
10	17	0	28	93	111	49
11	17	10	0	114	14	30
12	21	17	0	120	10	35
13	8	0	0	44	22	71
14	16	0	34	31	92	135
15	16	0	7	42	26	147
16	16	15	17	54	26	110
17	30	30	8	79	8	35
18	34	12	0	45	0	6
19	13	0	0	37	0	8
20	0	0	0	18	7	5
21	0	0	0	45	0	14
22	0	0	0	25	9	21
23	0	12	0	0	6	14
24	0	0	0	0	12	5
25	0	0	10	0	10	9
26	0	0	10	0	8	22
27	0	0	5	0	7	16
28	0	0	3	0	0	24
29	0	11	4	0	0	42
30	0	15	8	0	0	18
31	0	18	13	0	0	14
32	0	18	15	0	0	13
Base frequency	0.0269	0.0262	0.0305	0.0268	0.0302	0.0291

Note: small (within 10-30 percent) amplitude changes are permissible, which do not noticeably affect the sound. A simultaneous proportional change in all amplitudes of a certain sound is also possible - this corresponds to a change in volume. The data were obtained by averaging over 20 samples with a duration of 2-3 seconds each.

4. Synthesis of Artificial Words

When studying words, the problem of edge effects arises first of all. In the case of long vowels, this problem was easy to solve. The beginning and end of the recording were simply cut off by 10-30 percent of the total length of the recording, after cutting, there was a sufficient segment of the sound curve for studying

When studying words, you cannot do this, since during circumcision it was possible to accidentally cut off the sounds that are included in the word and are essential for its recognition

Therefore, the following decision was made: instead of a certain word, in one breath, a sequence of 3 of the same words was pronounced, forming a multiword in total. So, to study the word MALINA, the multiword

MALINAMALINAMALINA was written down (it is more convenient to denote it as 3MALINA), if possible, so that it was represented by one word for the speaker.

Then the base frequency of the multiword was determined and the sound curve was decomposed according to the proportional catching network [5]. After that, the drifting amplitudes were visualized and a search was made for recurring characteristic areas (notches), at least for one of the modes

In some cases, it turned out that it makes sense to perform some mathematical transformations in order to search for notches. So, for example, for the multiword 4MALINA it turned out that if we carry out the numerical differentiation of the amplitudes in time, then the notches are clearly visible in mode No. 3 (Figure 2).
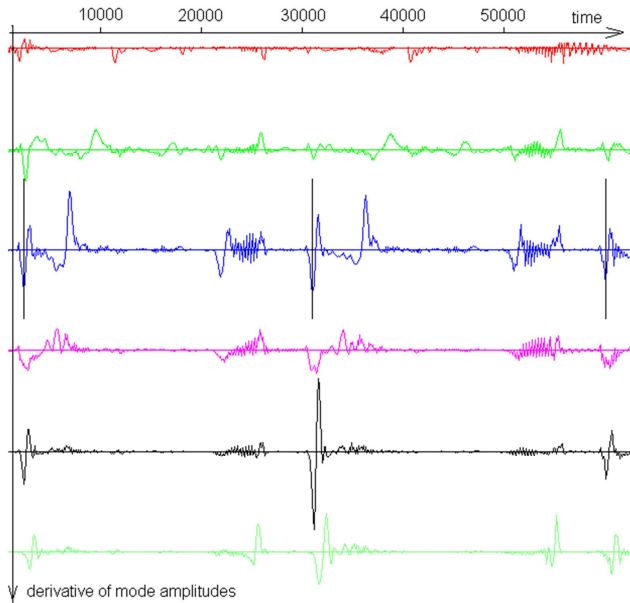
Figure 2. Graphs of the derivatives of the total amplitudes of the first 6 integer modes of one of the samples of the multiword 3MALINA in time. For ease of viewing, the graphs are spaced vertically, in accordance with the mode number, from top to bottom. Notches are marked on the graph of fashion # 3 (blue line) with vertical black straight lines intersecting the blue one.

As can be seen from Figure 2, the notches are also clearly visible for modes 5 and 6, albeit at points slightly shifted relative to the notches of mode No. 3 on the time axis. Obviously, in the interval between adjacent notches, each of the sounds of the word (phonemes) occurs exactly once, although the word does not necessarily begin with the notch.

Thus, everything that is between adjacent notches should be considered not as a word, but as a representative of the word. The word differs from the word representative in that some part of the word is cut off from its beginning and transferred to the end, as a result of which a word representative is created. For analysis, it is more convenient to use word representatives, which is done in this article.

Since in all samples a multiword consisted of 3 joined words, respectively, in each sample of a multiword 4 notches were searched for and, accordingly, 3 representatives of the word were selected. After the decomposition of the representative of the word into modes, the restoration of the representative of the word was carried out, several identical copies of which were successively docked with each other. As expected, the recovered multiword sounded the same as the original multiword.

To create those words that can be called artificial, one should find a certain mathematical formula, like formula (6), the use of which will allow generating a word. Below will be presented mathematical formulas obtained on the basis of one of the samples of representatives of the word MALINA

Finding these formulas began with the expansion of the sound curve of the multiword 4MALINA in a proportional catching network where number of modes is equal to 24 and at the base frequency $\omega_1 = 0.025$. Then, the notches were visually found and the representatives of the words were

distinguished, which were subjected to further study.

Since the word includes different sounds, and since the corresponding amplitudes of modes, in accordance with Table 1 - are different, then it will not work to approximate the amplitudes of the modes of the word representative with constant numbers. Therefore, a different path was chosen. As it turned out, for all modes, their amplitudes look like the sums of bell-shaped functions partially drifting over each other. (See Figure 3 below). Consequently, the amplitudes of the modes can be approximated by the sum of several Gaussian functions with different parameters

$$C_k(i) = \sum_{l=1}^{m} A_{k,l} \exp\left[-\frac{(i - \mu_{k,l})^2}{\sigma_{k,l}} \right] \quad i=1\ldots n, \; k=1\ldots 24 \quad (7)$$

Here i is the time (the number of the record count), C_k - is the drifting amplitude of the mode number k, m is the number of Gaussian functions approximating the amplitude of the mode number k, (in the present study, everywhere m = 6), $A_{k,l}, \mu_{k,l}, \sigma_{k,l}$ are the parameters of the Gaussian functions, n is the word length (in sampling counts, in the present study $n = 29299$).

These parameters can be selected by one of the known methods - for example, the coordinate-wise approximation method, the steepest descent method, or any other of the nonlinear approximation methods. Let us present the result of such an approximation using the example of the amplitudes of the first 8 modes of the representatives of the word MALINA (Figure 3).
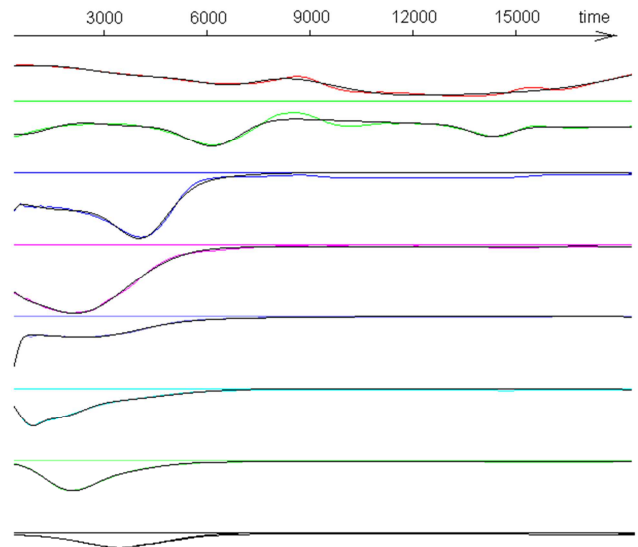
Figure 3. Amplitudes of the first 8 modes of the word MALINA representative, from top to bottom in numerical order. The colored curved line is the amplitude of the drifting mode, the flat horizontal line of the same color is its zero point. The black line superimposed on the colored line and partially covering it is the result of the approximation of the amplitude by the sum of the Gaussian functions. Word length is 29299 points.

As can be seen from Figure 3, there is some difference between the approximated and approximated functions. However, when you reversely synthesize a multiword, this difference is not audible.

Table 2. Parameters of Gaussian functions (7) approximating the amplitudes of modes 1-8.

		k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
l=1	μ_{kl}	16730	17893	4323	2354	2774	2145	22530	21836
	σ_{kl}	9721	13579	1530	1707	2384	1639	350	1305
	A_{kl}	424	349	1174	943	337	168	205	293
l=2	μ_{kl}	16849	6563	22228	22501	380	23359	2596	4208
	σ_{kl}	815	1432	1743	2002	250	591	1435	1965
	A_{kl}	157	627	911	319	425	208	420	134
l=3	μ_{kl}	13573	21594	1483	827	22480	1107	22082	20307
	σ_{kl}	3011	1231	1279	684	1870	410	1854	213
	A_{kl}	384	405	329	355	214	490	218	168
l=4	μ_{kl}	6006	14569	4854	23225	24021	23927	23285	5312
	σ_{kl}	6638	847	493	212	433	284	440	340
	A_{kl}	681	255	713	155	81	536	504	84
l=5	μ_{kl}	24635	647	279	4650	1030	21712	5313	3885
	σ_{kl}	198	471	286	1464	116	1733	953	326
	A_{kl}	135	270	469	211	56	162	120	69
l=6	μ_{kl}	-378	23894	23649	23884	622	3490	19962	21669
	σ_{kl}	1913	446	688	224	83	2242	233	661
	A_{kl}	98	207	409	145	59	171	161	166

Table 3. Parameters of Gaussian functions (7) approximating the amplitudes of modes 9-16.

		k=9	k=10	k=11	k=12	k=13	k=14	k=15	k=16
l=1	μ_{kl}	4954	15200	5119	13537	21521	14556	10661	11294
	σ_{kl}	1093	322	922	1386	4111	586	572	856
	A_{kl}	215	98	121	104	59	115	164	82
l=2	μ_{kl}	21888	6992	6721	5376	5298	11574	13438	13950
	σ_{kl}	1400	5854	9238	759	1723	1333	1106	1157
	A_{kl}	163	43	17	127	117	45	56	21
l=3	μ_{kl}	7889	22192	9970	22117	13747	23722	5219	21801
	σ_{kl}	1079	1783	294	1547	2486	10699	2266	1931
	A_{kl}	94	52	99	61	61	23	25	14
l=4	μ_{kl}	5338	5452	22145	13225	9907	9927	22018	6065
	σ_{kl}	468	654	2018	287	452	321	4521	9833
	A_{kl}	451	48	48	39	59	50	18	7
l=5	μ_{kl}	15522	10650	11319	14229	21835	13606	11831	4696
	σ_{kl}	383	781	961	658	1086	208	391	1292
	A_{kl}	68	30	91	156	115	37	61	10
l=6	μ_{kl}	6301	21104	14450	5738	23694	5702	13329	21273
	σ_{kl}	299	693	1095	11539	146	690	112	351
	A_{kl}	121	46	58	22	28	15	22	8

Table 4. Parameters of Gaussian functions (7) approximating the amplitudes of modes 17-24.

		k=17	k=18	k=19	k=20	k=21	k=22	k=23	k=24
l=1	μ_{kl}	11001	10789	11399	22447	2957	7025	6663	5063
	σ_{kl}	895	551	1552	1584	2224	14510	13426	721
	A_{kl}	57	41	99	58	22	4	3	23
l=2	μ_{kl}	23261	21952	22397	10229	11011	4883	9687	5555
	σ_{kl}	853	1969	1744	5447	6237	275	213	211
	A_{kl}	18	31	40	24	9	7	8	19
l=3	μ_{kl}	13741	4245	3968	5305	22491	9632	5377	3792
	σ_{kl}	1698	3442	3829	569	2104	226	279	412
	A_{kl}	20	20	35	57	10	12	3	13
l=4	μ_{kl}	4561	11603	21250	12022	3705	3806	12670	8576
	σ_{kl}	3661	412	530	726	209	1759	251	13198
	A_{kl}	16	41	73	35	16	7	4	4
l=5	μ_{kl}	20353	12337	23905	1647	1398	8543	14156	1789
	σ_{kl}	256	3179	414	1451	283	296	358	243
	A_{kl}	19	25	37	48	13	6	3	6
l=6	μ_{kl}	21393	20473	22888	3798	10113	7368	11834	22493
	σ_{kl}	726	200	204	461	389	381	227	901
	A_{kl}	30	17	42	46	11	7	3	4

Thus, the drifting amplitudes of word representatives can be represented in the form (7). The necessary values of the parameters of the Gaussian functions for the amplitudes of the modes of the word MALINA are given in Tables 2-4.

The number k in the tables 2-4 means the number of the mode, the number l is the number of the Gaussian function included in the approximating sum (7).

As for the phases, their typical behavior is presented below (Figure 4).
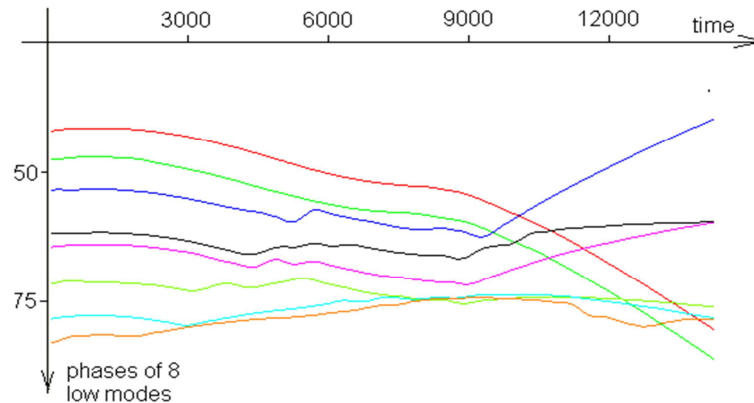
Figure 4. *Normalized phases of the first 8 modes of one of the representatives of the word MALINA (Normalized phase is the phase divided by the mode number 1). The first fashion at time=1 is red, the second is green, the third is blue, the fourth is purple, the fifth is black, the sixth is turquoise, the seventh is blue, the eighth is brick.*

As can be seen from Figure 4, in some cases, the phases experience a sharp break. Such functions are inconvenient for approximation by power functions, therefore, it was decided to split the entire length of a word representative into sections, and as the boundaries between these sections, select those moments in time where at least one of the phases experiences a break.

Between the boundaries of the sections, the phases can be linearly interpolated. The boundaries of the sections of linear interpolation and the phase values at the boundary points for modes 1-24 are given in Tables 5-7.

Tables 5-7 X-coordinates of the boundary points of the splitting sections of the representative of the word MALINA are given in the first column. The number k in these tables denotes the mode number.

Table 5. *Boundaries of sections and phase values at boundary points for modes 1-8.*

i	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
1	0.95	3.09	-50.42	-79.36	-159.87	20.04	-102.84	-69.90
690	-0.82	-0.32	-53.68	-86.03	-166.07	13.17	-112.42	-69.32
1290	-1.95	-2.64	-57.09	-90.63	-172.00	8.63	-119.83	-82.09
2700	-2.86	-3.44	-59.62	-94.68	-177.02	3.75	-125.56	-89.04
3390	-2.46	-2.29	-58.44	-93.57	-175.31	6.15	-122.18	-85.93
4590	-0.66	1.55	-53.65	-86.50	-166.43	16.94	-108.39	-99.99
5100	0.43	3.82	-51.29	-82.91	-160.86	22.05	-113.91	-105.02
6330	3.37	9.17	-42.68	-88.78	-161.90	13.28	-126.01	-113.96
7200	5.27	12.25	-36.78	-81.00	-157.86	16.27	-124.09	-122.30
8640	8.67	18.36	-25.95	-81.88	-143.67	11.49	-138.23	-132.82
9420	10.15	21.75	-21.34	-78.87	-140.03	22.89	-147.16	-139.02
9840	11.08	23.42	-24.56	-76.15	-137.24	27.44	-148.26	-134.91
10200	12.06	25.03	-22.28	-79.62	-134.69	29.55	-151.03	-136.51
12000	17.33	35.56	-6.47	-108.76	-153.05	11.13	-164.12	-144.46
13050	19.94	40.81	1.70	-120.18	-149.94	11.30	-174.91	-153.02
14100	22.41	45.86	8.73	-119.74	-138.03	5.91	-186.72	-162.71
15600	24.75	50.67	16.90	-109.56	-125.63	21.17	-191.30	-177.79
15990	24.63	49.94	13.26	-108.77	-123.90	22.69	-190.08	-177.05
18600	21.37	43.47	3.62	-120.96	-137.45	34.35	-188.69	-184.76
20550	14.43	30.03	-14.32	-107.43	-154.79	31.90	-189.90	-199.34
21030	12.47	26.16	-12.31	-112.03	-154.90	29.84	-184.95	-196.84
22530	5.56	13.18	-1.39	-108.91	-157.91	22.02	-172.21	-191.61
24180	-2.59	-2.91	7.84	-106.53	-164.44	23.60	-178.08	-203.46
24720	1.36	-9.13	5.98	-106.10	-168.19	21.67	-179.16	-207.24
26070	-4.77	-21.34	13.36	-100.27	-174.73	30.70	-172.06	-211.59
29299	-14.84	-41.38	-15.57	-79.23	-213.37	33.12	-162.31	-227.07

Table 6. Boundaries of sections and phase values at boundary points for modes 9-16.

i	k=9	k=10	k=11	k=12	k=13	k=14	k=15	k=16
1	-92.32	16.92	39.61	97.73	69.52	-25.98	20.94	66.23
690	-94.15	14.08	38.91	96.60	71.09	-23.75	21.16	55.62
1290	-90.37	16.36	41.01	96.95	70.02	-26.78	32.21	60.45
2700	-95.54	8.80	33.78	99.68	63.88	-25.78	74.95	59.93
3390	-93.66	12.37	30.53	103.55	65.19	-18.26	79.34	61.52
4590	-107.07	12.51	19.12	102.16	56.60	1.04	77.11	58.67
5100	-110.23	9.75	17.81	101.71	57.36	3.60	79.86	61.54
6330	-115.97	6.94	17.68	105.23	62.37	17.89	90.07	63.83
7200	-122.30	2.14	15.88	106.60	62.89	19.71	96.10	72.20
8640	-128.94	-0.82	15.96	107.97	68.99	28.68	100.70	70.22
9420	-130.67	-4.61	13.61	107.98	73.08	32.71	108.13	74.33
9840	-132.28	-6.98	10.53	104.97	73.94	29.25	109.08	78.31
10200	-134.09	-7.00	10.31	106.36	75.45	30.39	106.68	75.77
12000	-136.47	-4.13	21.38	119.18	95.21	51.75	100.23	61.34
13050	-142.32	-7.20	20.59	122.07	100.95	58.69	111.28	61.51
14100	-149.07	-12.20	18.20	122.29	102.97	63.19	117.27	65.99
15600	-145.22	-0.29	7.60	115.94	96.14	54.98	121.72	71.91
15990	-144.82	2.14	10.01	118.42	99.13	53.86	126.75	88.93
18600	-149.07	13.85	12.70	123.59	107.83	47.48	125.86	122.16
20550	-155.03	18.73	15.98	126.84	108.99	56.81	127.40	237.54
21030	-152.85	18.49	13.69	125.05	113.08	50.36	127.90	236.43
22530	-153.01	13.90	2.66	135.48	116.89	47.44	120.85	220.89
24180	-156.85	19.39	-0.02	131.12	117.59	42.29	115.12	211.79
24720	-154.47	21.82	-0.76	133.71	116.95	43.36	114.07	205.35
26070	-154.04	26.22	4.62	134.15	111.33	39.10	114.49	218.35
29299	-164.19	18.14	17.06	149.54	109.52	43.21	113.47	209.18

Table 7. Boundaries of sections and phase values at boundary points for modes 17-24.

i	k=17	k=18	k=19	k=20	k=21	k=22	k=23	k=24
1	311.71	169.99	63.09	227.08	526.45	768.11	2296.48	2576.68
690	314.96	170.97	70.72	221.12	522.02	833.38	2408.22	2630.22
1290	313.03	174.25	73.20	222.69	515.66	848.96	2460.27	2647.50
2700	310.35	177.35	74.92	223.61	510.13	841.25	2618.07	2683.95
3390	316.31	178.84	77.35	226.18	519.47	837.06	2666.77	2717.94
4590	317.54	179.83	79.75	229.42	519.95	841.05	2879.09	2734.03
5100	321.17	185.15	85.74	225.66	514.90	837.70	2973.62	2739.71
6330	322.84	182.18	76.59	219.42	530.76	862.17	3068.99	2744.20
7200	337.69	192.95	72.04	209.48	550.89	881.98	3096.14	2741.58
8640	344.20	184.58	61.81	203.30	564.57	893.97	3115.68	2786.30
9420	351.35	179.09	58.09	196.83	564.57	891.53	3122.48	2835.03
9840	350.60	180.05	60.11	196.94	563.20	889.42	3119.50	2833.54
10200	347.37	179.22	59.97	197.40	563.21	909.04	3142.71	2834.43
12000	337.43	174.29	60.18	203.58	577.18	1048.50	3272.57	3064.90
13050	349.22	162.46	52.16	198.51	589.21	1119.77	3284.78	3211.02
14100	360.48	164.60	43.24	190.17	590.02	1152.99	3391.66	3292.31
15600	362.49	169.46	48.20	198.29	604.64	1217.41	3533.45	3523.92
15990	374.23	167.94	78.91	213.47	607.16	1309.90	3545.69	3526.30
18600	446.67	166.39	107.26	276.56	845.99	1377.37	3798.80	3770.87
20550	516.49	181.20	157.22	349.10	985.36	1720.73	4068.01	3976.55
21030	517.00	178.78	160.18	348.40	1001.69	1767.36	4115.29	4034.23
22530	524.84	181.85	156.67	338.35	1028.70	1864.55	4308.88	4169.39
24180	528.21	185.49	156.04	330.36	1016.32	1946.68	4486.44	4257.09
24720	528.74	203.07	167.02	333.34	1012.97	1976.87	4534.50	4376.86
26070	563.06	213.81	171.87	356.10	1021.57	2163.41	4688.76	4581.78
29299	752.59	310.47	253.59	460.31	1177.22	2282.86	4856.29	4771.62

5. Conclusion

The amplitudes of each of the modes can be approximated by the sum of Gaussian functions, the phases can be approximated piecewise linearly. After such an approximation, itis possible, according to formula (4), to recreate a word representative, sequentially join anynumber of identical word representatives, and thereby obtain a reconstructed multiword, which can be converted into any of the audio formats and listened to.

Note. The numbers in columns 2-9 of tables 5-7 are given with an accuracy of 0.01. These numbers can be multiplied by 100, added to them the numbers from tables 2-4 and the

first column of tables 5-7, the result is a set of 1082 integers, which occupy 4328 bytes in computer memory. The word MALINA written in. Wav format is about 59600 bytes. Thus, the content of tables 2-7 can be considered as a result of compression of the word MALINA, the compression ratio is about 14. (Moreover, it can be increased due to more economical methods of writing numbers, discarding low-intensity modes, etc.).

References

[1]   Auni Hannum. Speech Recognition is not Solved. 2017, Posted on October 11, 2017.

[2]   1. Sorokin, V. N. Speech recognition based on spectral-temporal irregularities in the speech signal. Acoustic magazine, 2020, V66, N1, 71-85.

[3]   Yakovlev, A. V., Sosnin, V. A. Digital processing of acoustic pulses in the acoustic emission diagnostics system KAEMS, 2018, N3, http:/ejta.org, 2018.

[4]   Vasilieva, L. G., Zhileikin, Ya. M., Osipik Yu. I. Fourier transforms and wavelet transforms. Their properties and applications. // Computational methods and programming: in 3 volumes - M., - V 3, - Issue 1, - P 172-175, 2002.

[5]   Maksimchuk, I. V., Gergel, L. G., Osadchiy, O. V. Comparative analysis of Fourier and wavelet transform for the analysis of the photoplethysmogram signal. [Electronic resource] // Modern scientific research and innovation - M., 2013. - No. 6.

[6]   Mitsianok, V. V. Determination of the numerical characteristics of high-frequency speech sounds based on approximation by harmonic functions // Bulletin of the National Academy of Sciences of Belarus, ser. f.-m.n., - Minsk, - No. 2, P. 111-118. 2009.

[7]   Mitsianok, V. V. On the physical structure of simple vowel sounds of human speech // Open semantic technologies for the design of intelligent systems: materials of the VI international scientific and technical conference OSTIS-2016, Minsk, February 18-20, 2016, -Minsk: BSUIR, 2016, p. 404-410.

[8]   Lobanov, B. M., Galunov, B. I., Zagoruiko, N. G, Ontology of the Subject domain "Speech Signal Recognition and Synthesis" // Procttding of the international conference "Speech end Computer" St.-Petersburg, 2004. – 440-444.

[9]   Lobanov, B. M., Solomennik, A. I., Zhitko, V. A. An experience of an objective assessment of the intonation quality of synthesized Russian speech. Computational linguistics and intelligent technologies. Based on the materials of the conference "Dialogue" Moscow 2018, issue 17 (24). Publishing house of the Russian State University for the Humanities.

[10] Rusak, V. P., Getsevich, Yu. S., Mandric, V. A. Problems of norms, culture of language and speech generation. Collection of papers and abstracts of the 8th conference "Traditions and the current state of culture and arts. Minsk, Belarus. Minsk, Law and Economics, 2018, 748-752SPb., - 2013. No. 4.- Access mode: http://www.ejta.org, free.

[11] Sorokin, V. N., Viyugin, V. V. Tananikin, A. A. Personality recognition by voice: an analytical review. Information Processes, 2012. -t 12 - N. 1-30.

[12] Mitsianok, V. V. On the problem of identification and verification of personality by phase characteristics of speech sounds [Electronic resource] // Technical acoustics. - Electron. magazine - SPb., - 2015.- No. 7.- Access mode: http://www.ejta.org, free.

[13] Mitsianok, V. V. Generation of artificial sounds and words of human speech. Thesis of Annual Scientific Conference of Polessian Univ. 2021. Polessian Univ. Edition.

[14] Mitsianok, V. V. On the synthesis of artificial sounds of human speech sounds [Electronic resource] // Technical acoustics. - Electron. magazine - SPb., - 2017.- No. 1.- Access mode: http://www.ejta.org, free.

[15] Mitsianok, V. V., Konovalova, N. V. Application of phase analysis of speech sounds for recognizing a person by his voice. [Electronic resource] // Technical acoustics. - Electron. magazine - SPb., - 2013. No. 4.- Access mode: http://www.ejta.org, free.

[16] Mitsianok, V. V. On the physical structure of sounds Z, Zb, ZH, ZHb. [Electronic resource] // Technical acoustics. - Electron. magazine - SPb., - 2014.- No. 9.- Access mode: http://www.ejta.org, free.