V.I. Vorobev A.G. Davydov

SYNTHESIS OF SPEECH-LIKE SIGNALS

Belarus State Information and Radio Technical University
6 P.Brovki str., 220660 Minsk Belarus
Ph. (017) 239-8939, Fax: (017) 231-0914
E-mail: nil53@bsuir.edu.by

Among activities on speech synthesis the reports about so-called " Speech-Like Signals " (SLS) have appeared recently. Main areas, in which application of these signals is expected, are marked psychology and information protection [1,2].

Under already used in the indicated areas of SLS on a tacit consent the signals, phoneticsly similar to speech, but not carriers of semantic information accessible to traditionally used simple means (first of all - at direct listening of these signals by the subscriber) are perceived.

In psychologic experiments SLS reshape, in particular, by frequency inverse of a spectrum of customary signals of speech. At protection of information with the help SLS the latter are obtained by summation of three arbitrary voice calls [3].

Both in the first and in the second cases the term " Speech-Like Signals " seems of to be not quite correct, since the reshaped signals contain in both cases in a latent kind of the semantic information of original speech signals. At proper processing of these "SLS" the information can be extracted. The cases are known, when the operators perceived the main sense of a speech inverted on spectrum, obviously not converting listened audio signals in initial voice. At, latent from consciousness, levels the operator irrespective of his desire inverted the spectrum of adopted signals and extracted the semantic contents of initial speech in consciousness from the obtained material.

It is deemed reasonable to attach an independent sence to the term " Speech-Like Signals ", reqardless of ways of signals generating and processing. For this purpose we shall assume these signals to be audio signals, temporary and spectral reflectance, and as to perception on audition recalling speech, but not bearing any semantic information.

As the main features of resemblance SLS also by likeness taking into account data available [4 - 7] about representative parameters of speech signals, it is offered to consider the following:

1. Bursted nature of temporary changes SLS with duration of splashes (15 - 120) milisec;
2. Grouping splashes in members on (2-8) pieces in each, with intervals between separate splashes (10 - 50) milisec, at an interval between members from 10 milisec up to 1 sec;
3. Availability in a spectrum SLS of a fundamental component with frequency (75 - 300) Hz;
4. Availability for SLS an envelope with a spectrum, concentrated in frequency band (9-60) Hz;
5. The form of an envelope bursted of a signal can be preset analytically with a variation of parameters or is incidentally selected from a set of the previously prepared signals;
6. Availability in a spectrum from three up to five formants, which one can be formed under the law:

$$f_n(t) = A_n e^{-\frac{t}{t_n}} \sin(\omega_n t - \varphi_n),$$

where: A_n - amplitude of n formant ($n = \overline{1,N}$); N - total number of formants; t - current time of readout, which one changes from a beginning of the season of a fundamental component and up to its end; t_n - time period, during which one the module of amplitude decreases in e of time (duration of this interval lies in limits (1,5 - 30) milisec); j_n - initial phase of harmonic carrier of n formant; w_n - circular frequency of formant ($w_n = 2pf_n$). The cyclical frequency of formants receives values in following indicated approximately [4, 6, 7] ranges: $f_1 = (300 - 1000)Hz$, $f_2 = (1000 - 2000)Hz$, $f_3 = (2000 - 3000)Hz$, $f_4 = (3000 - 4000)Hz$, $f_5 = (4000 - 5000)Hz$.

It is possible to note, that the data, borrowed from the references, about parameters of splashes and their members enter some inconsistency with reduced numeric values of temporary and frequency response curves of an envelope SLS. However, it can not serve an interrupting for further consideration.

At formation of SLS on the basis of usage of the mentioned above tags it is necessary to be guided by following padding reasons:

- The reshaped SLS implementation is expedient to divide into segments with duration from 5 up to 15 terms of a fundamental component. It is necessary to ensure, that on these segments the initial phases and the frequencies of carriers of oscillations of formants changed no more than on 10 % of the initial values. The nature of change of initial phases and frequencies of formants can be the same as well as change of an envelope of splashes;
- At transition from one segment to other frequency of a fundamental component, frequency of formants, amplitude, decay time of signals of formants can be changed by a jump on (20-70) % in any direction;
- At transition from one splash to other its parameters can be changed arbitrary in the limits which have been mentioned above.

For formation SLS it is possible to use two approaches:

- At a signal level, i.e. reshaping directly some process with temporary and spectral parameters adequate the offered above requirements to SLS;
- At a text level, i.e. generating of the text on some specially designed algorithms. The one, make sound by the means, specially intended for it, give SLS.

The second approach ensuring on this research stage a comparative simplicity of simulation studied SLS on a computer is studied. Below pursuant to this approach some ways of formation SLS can be offered.

Most simple is the formation of "words" with usage of absolute probabilities of occurrence of characters and gaps in the texts of modelled language (Table 1).

Table 1.

Absolute probabilities of characters of Russian alphabet.

Space	0,175	**Ê**	0,028	**×**	0,012
Î	0,090	**Ì**	0,026	**É**	0,010
Å	0,072	**Ä**	0,025	**Õ**	0,009
À	0,062	**Ï**	0,023	**Æ**	0,007
È	0,062	**Ó**	0,021	**Þ**	0,006
Ò	0,053	**ß**	0,018	**Ø**	0,006
Í	0,053	**Û**	0,016	**Ö**	0,004
Ñ	0,045	**Ç**	0,016	**Ù**	0,003
Ð	0,040	**Ü,Ú**	0,014	**Ý**	0,002
Â	0,038	**Á**	0,014	**Ô**	0,002
Ë	0,035	**Ã**	0,013		

Advantage of this way, is the automatic partitioning of the reshaped text on words by gaps appearing with probability 0,175.

The example of the text obtained by this way looks like: " *îìòãáóñï ãàäíí ààá ð ÷éäâäíóãåéíáîìòèþ à ò èðåïðí ï êñ ñõëòèåêèóÿòèîàä ðë àþàñ ëêï îíåá ï ðóí÷ù üèýü òê íêíéáåáîò åìÿâùòí ìà ãóèîòäåñàä à òâèñüëñèíá éìîòòåàäçî åøåäåëà èâàóà ìëèö ëï àÿíóàóîò ðáèóò îáíå ðå îâååï êàìîîîäîç ûå îíäóñîîþðåðî ÷ìóìëêêøåóñîûîéï* ".

Another way of formations SLS is the synthesis of words on the basis of probabilities of occurrence of different syllables. Thus the distinction of probabilities of occurrence of the given syllable is allowed at different quantities precursor to him in a word of other syllables. For allocation alone of syllables it is expedient to use following assumptions: a) is supposed that one vocal character, is in one syllable; b) the syllable can consist only of one vocal character; c) the syllable endes in the vocal character; d) characters at the end of the words which have stayed after formation of syllables with allowance of maiden three assumptions, are considered as a separate syllable.

The listed assumptions visually are illustrated by an example.

Initial word: “Ï ÎËÓÏ ÐÎÂÎÄÍÈÊ”.

Exact splitting: “Ï Î-ËÓ-Ï ÐÎ-ÂÎÄ-ÍÈÊ”.

Splitting obtained on the basis of reduced assumptions: “Ï Î-ËÓ-Ï ÐÎ-ÂÎ-ÄÍÈ-Ê”.

As it is visible, in outcome the precise splitting receives practically. The errors are allowed only at formation of syllables of “ÂÎÄ” and “ÍÈÊ”. However, even at such splitting the obtained members of a word prolong "to sound".

The "syllable" “Ê” has appeared at the end of a word. It is an incorrect syllable. It is interesting, that it can meet, only at the end of a word and with small probability. In this connection, it is necessary to enter correction: all syllables, general probability of occurrence which one less defined value are rejected as non valid.

Example of the text, formed by the second way is placed below: " *òèåáùåëÿ ï îçîâùà äìíèâàäå ôàï åàÿëüíî ï åäîñòàåãîìû âñåðîêòèå èçäåðüå ýíå êî ðàñòîãàéå ìà îñîêî íåçíèäíûíîííûò ññòüåûé÷íî èêäåé ìíîêðîðàéâ âîðûé ÷òîï îñòðîé èäíûñíåÿòñÿ ï äï ãóõéñàòå ó òðå àñï îçóòüíòíûÿ ëèñòðîëüçóâà ñâÿðàíèíèâàòñÿ âáùå çîëåâíþà ï îêòîáðàíè íàëüøåõîíòàî íåòèëíÿåëü î òèìÿ÷êèòêèþì àòðèðîâÿ òèíôèáðàâî âûìÿôèíèáðà äëÿëÿîó äàï ðè äâàòîêìõííî*".

For the purposes of formation SLS the definite concern introduces also compiling of the texts, on the basis of usage of probabilities of occurrence of different notes in language. The values of probabilities of occurrence of definite notes in discharged parts of words are indicated in Table. 2.

Table 2

Occurrence of notes in %.

Agree and their groups								vocal (impact) in middle of a word
Initial				final				
Á 3,7 Â 7,3 Ã 2,0 Ä 4,8 Æ 0,8 Ç 3,9 Ê 4,9 Ë 1,6 Ì 4,8 Í 12,6	Ï 11,6 Ð 3,6 Ñ 8,4 Ò 6,4 Ô 2,1 Õ 1,0 Ö 0,3 × 1,3 Ø 1,1 Ù 0,1	ÂÍ 0,3 ÃË 0,2 ÃÐ 0,1 ÄÂ 0,3 ÄË 0,4 ÄÐ 0,3 ÇÄ 0,4 ÇÍ 0,6 ÊÂ 0,1 ÊÐ 0,3	Ï Ë 0,3 Ï Ð 3,7 ÑÂ 0,7 ÑÊ 0,4 ÑÏ 0,2 ÑÒ 1,1 ÒÐ 0,6 ÔÑ 1,1 ×Ò 0,9 ØÒ 0,9	É 17,0 Ê 5,5 Ë 7,7 ËÜ 0,4 Ì 14,1 Í 5,1 ÍÜ 0,8 Ï 0,2 Ð 1,3 ÐÜ 0,5	Ñ 2,9 ÑÜ 3,5 Ò 12,8 ÒÜ 9,3 Ô 4,8 Õ 8,4 Ö 0,4 × 0,8 Ø 2,9 Ù 0,3	ÇÌ 0,1 ÇÍÜ 0,2 ÊÏ 0,1 ÊÒ 0,3 ÑÊ 0,1 ÑÒ 0,2 ÑÒÜ 1,1 ÐÊ 0,03 ÐÏ 0,03 ÔÊ 0,03	Ì Ü 0,08 Ï Ü 0,03 ÔÜ 0,03 Ï ÜÔ× 0,05	À 32,5 Å 19,6 ¨ 1,3 È 14,3 Î 14,6 Ó 7,0 Û 6,1 Ý 1,2 Þ 0,6 ß 2,8
In the beginning of words å ¸, þ ÿ– 4,8								

For lack of a place algorithmic details of formation of the text the third way here are not discussed.

The example of the formed text looks like: " *ï óçäàòó ï àíîôàêû íàñèõó ï åàñàìàòàé ï àçåæèô ñàï ðûï àíîøà íèì äëèêàõàô øÿï àíûôåõÿ ñèòüÿ âàï àæèõÿòà ðîìàõîìó øò ñòåæèòîï àñüÿòè íàï ðîòà íÿâåðî ñî ôñàòî* *âóâ¸êèéœòå ï îíàñœòü ï œâîéœè ðàð ï èíœœèêîì ëàñâàéîé ñàì œõàé ãàì ûíèì ý òó ÷èøòèñàòüèì òàì ûêàêóô êœï îò å áîíîõà ê ñîâíœòàø ñûíü â*".

After formation of sequences of notes or characters got by the one of it is necessary to receive analogue representation of SLS. For this purpose in modern computers there is a number methods and algorithms [7 - 10], distinguished by complexity and requirements to a computer storage.

It is possible to receive a subjective estimation of examples of the texts formed for generation SLS, by the reading them aloud. The comparative analysis of different ways of synthesis SLS leaves for frameworks of the report. The state of the art of mining of problems, formation and usage of Speech-Like Signals having probabilistic nature, testifies that they can find application for protection of voice informations in communication channels by masking intervals between transmission sessions, in quality noise protective of signals in systems of audio protection of the information from outflow on acoustic channels, and also for special noise protective systems of security and fire signallings. Area of possible application SLS is also psychology.

R E F E R E N C E S

1. Davydov G.V., Potopovich A.V. Protection of a voice information noise speech-like signal.//Informations of the Byelorussian engineering academy. ¹ 1(9)/1, 2000., p 123-125. (In Russian).
2. Physiological soundman. // http://psi.webzone.ru/st/005300.htm
3. Kalinin S.V. About some new traditions in development of systems chatter-acoustic noising. //http://www.MASCOM.ru/vib_st.html.
4. Sapojkov M.A. Voice call in cybernetics and communication (transformation of speech, with reference to problems of engineering and cybernetics) Sviazizdat , 1963. – 452 p. (In Russian).
5. Sapojkov M.A., Michailov V.G. Vocoder communication. Wireless and communication 1983 248 p (In Russian).
6. Markel Dj., Grey A.H. Linear prediction of speech: translation with English. /By edition. J.N. Prohorova and V.S. Zvezdina Communication 1980. 380 p. (In Russian).
7. Petluchenko N.V. The phonetic characteristics some agree in the german prepared speech. // Transactionses XI of session of the Russian acoustic company. Modern voice technologies : GEOS 1999. p 112 – 117. (http://www.akin.ru/ras.htm).
8. Brovchenko V.A., Voloshin V.G., Chernov O.F. Computer learning systems of intonation of speech. // Transactionses XI of session of the Russian acoustic company. Modern voice technologies : GEOS 1999. p 108 – 111. (http://www.akin.ru/ras.htm).
9. Rumianzev M.K., Kaplun M.I., Pospelov B.V., Dolotin K.I. Linguistic aspects of speech synthesis. // Transactionses XI of session of the Russian acoustic company. Modern voice technologies : GEOS 1999. p 103 – 104. (http://www.akin.ru/ras.htm).
10. Krivnova O.F., Zaharov L.M., Zinoveva N.V., Strokon G.S., Babkin A.V. Experience of system development of automatic speech synthesis for Russian. // Transactionses XI of session of the Russian acoustic company. Modern voice technologies : GEOS 1999. p 93 – 99. (http://www.akin.ru/ras.htm).