Праведзены спектральны аналіз (STFT, мел-спектраграмы, MCD = 5,2 ± 0,4 дБ) і ацэнка прадукцыйнасці (RTF = 0,31 з Numba JIT) пацвярджаюць, што рэалізаваная сістэма забяспечвае высокую якасць сінтэзу гаворкі ў рэальным часе. Верыфікацыя на аснове параўнання асцылаграм і інтанаграм паказала карэктную перадачу суперсегментных прасадычных характарыстык, а ацэнка апаратнай рэалізацыі на платформе FPGA класа Artix-7 (~35 % рэсурсаў) пацвердзіла тэхнічную магчымасць пераносу вакодэра на ўбудаваныя сістэмы.

Літаратура

1. Kubichek, R. Mel-cepstral distance measure for objective speech quality assessment / R. Kubichek // Proceedings of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing. – 1993. – P. 125–128.
2. Markel, J. D. Linear Prediction of Speech / J. D. Markel, A. H. Gray. – Berlin : Springer-Verlag, 1976. – 288 p.
3. Morise, M. WORLD: a vocoder-based high-quality speech synthesis system for real-time applications / M. Morise, F. Yokomori, K. Ozawa // IEICE Transactions on Information and Systems. – 2016. – Vol. E99-D, N 7. – P. 1877–1884.
4. Выгонная, Л. Ц. Інтанацыя. Націск. Арфаэпія / Л. Ц. Выгонная. – Мінск : Навука і тэхніка, 1991. – 215 с.
5. Xilinx Inc. Artix-7 FPGA Family Data Sheet: DC and AC Switching Characteristics. – San Jose: Xilinx, 2020.
6. Рабинер, Л. Теория и применение цифровой обработки сигналов : пер. с англ. / Л. Рабинер, Б. Голд ; под ред. Ю. Н. Александрова. – М. : Мир, 1978. – 848 с.

## ПРАГРАМНАЯ РЭАЛІЗАЦЫЯ ВАКОДЭРА І МІКШЫРАВАННЕ СЕГМЕНТНЫХ І СУПЕРСЕГМЕНТНЫХ КАМПАНЕНТАЎ ПРЫ СІНТЭЗЕ МАЎЛЕННЯ

**С. Ю. Крышнеў**

*Установа адукацыі «Гомельскі дзяржаўны ўніверсітэт імя Францыска Скарыны», Рэспубліка Беларусь*

**І. У. Блыхін**

*Установа адукацыі «Гомельскі дзяржаўны тэхнічны ўніверсітэт імя П. В. Сухога», Рэспубліка Беларусь*

Навуковыя кіраўнікі: Ю. В. Крышнеў, М. У. Буракова

*Установа адукацыі «Гомельскі дзяржаўны тэхнічны ўніверсітэт імя П. В. Сухога», Рэспубліка Беларусь*

*Разгледжана праграмная рэалізацыя вакодэра сістэмы сінтэзу маўленчых акустычных сігналаў, якая забяспечвае аднаўленне гукавога сігналу па мел-кепстральных каэфіцыентах з улікам кіравання суперсегментнымі (прасадычнымі) параметрамі выказвання – контурам асноўнага тону, інтэнсіўнасцю і працягласцю. Апісаны прынцыпы мікшыравання сегментнага і суперсегментнага патокаў даных пры сінтэзе гаворкі ў рэальным часе.*

**Ключавыя словы:** сінтэз маўлення, вакодэр, мел-кепстральныя каэфіцыенты, прасадычная мадуляцыя, PSOLA, мікшыраванне, суперсегментны ўзровень, WASAPI.

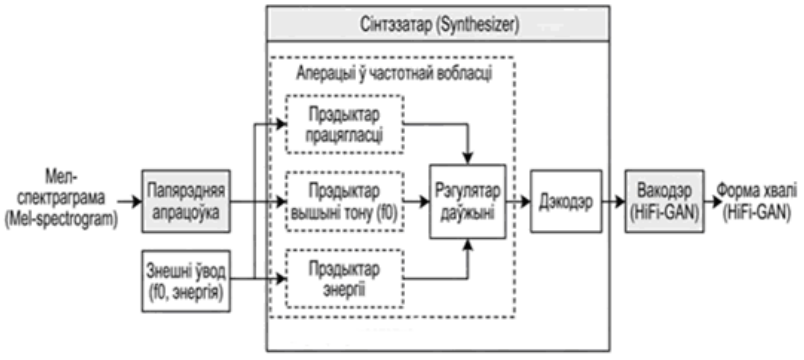
Сучасныя сістэмы сінтэзу маўлення (text-to-speech – TTS) пераўтварыліся з дапаможнага сродку агучвання тэксту ў неад’емны кампанент галасавых памочнікаў, аўтаматычных даведачных сістэм і сродкаў агучвання інфармацыі для людзей з парушэннямі зроку. Пры гэтым якасць успрымання сінтэзаванай гаворкі вызначаецца

не толькі пераборлівасцю асобных гукаў, але і натуральнасцю інтанацыйнага малюнку выказвання, які фарміруецца сукупнасцю суперсегментных (прасадычных) параметраў – частатой асноўнага тону, інтэнсіўнасцю і працягласцю складаў.

Прадметам праекта, на матэрыялах якога падрыхтаваны дадзены даклад, з'яўляюцца метады і алгарытмы лічбавага сінтэзу гаворкі з кіраваннем прасадычнымі характарыстыкамі, а мэтай – распрацоўка праграмна-апаратнага комплексу для сінтэзу натуральнай гаворкі з магчымасцю дынамічнага кіравання інтанацыйным малюнкам выказвання. Цэнтральным звяном такой сістэмы з'яўляецца вакодэр, і менавіта яго праграмнай рэалізацыі і механізму мікшыравання сегментных і суперсегментных кампанентаў прысвечана дадзеная праца.

**Агульная архітэктура праграмнага вакодэра.** Вакодэр (ад англ. «voice coder») з'яўляецца цэнтральным звяном сістэмы сінтэзу гаворкі: ен пераўтварае набор акустычных параметраў (мел-кепстральных каэфіцыентаў, контуру асноўнага тону $f_0$, параметраў аперыядычнасці) у гукавы сігнал часавай вобласці, прыдатны для прайгравання праз гукавую карту. У праграмнай рэалізацыі, арыентаванай на настольны камп'ютар, вакодэр працуе з буферамі PCM-даных (лічбавы гукавы паток фармату Pulse Code Modulation), якія перадаюцца аперацыйнай сістэмай у аўдыедрайвер праз адпаведны API.

У рабоце рэалізаваны вакодэр тыпу WORLD [1], адаптаваны для працы ў рэжыме сінтэзу з падтрымкай відавочнага кіравання асноўным тонам $f_0$ і параметрамі спектральнай ахінальнай. Архітэктурна сістэма падзяляецца на пяць модуляў (мал. 1): блок разліку спектральнай ахінальнай па мел-кепстральных каэфіцыентах; блок сінтэзу сігналу ўзбуджэння з кіраваннем па $f_0(t)$ і аперыядычнасці BAP(t); блок суперпазіцыі (фільтраванне ў частотнай вобласці); блок прасадычнай мадуляцыі, які накладае суперсегментныя элементы – контур $f_0$, ахінальную інтэнсіўнасці $E(t)$ і паўзы; блок мікшавання і вываду, які запісвае PCM-даныя ў кальцавы буфер і перадае іх на гукавую карту.




*Мал. 1*. Функцыянальная схема праграмнага вакодэра

Прынцыпова важна адрозніваць два ўзроўні апрацоўкі сігналу. Сегментны ўзровень адказвае за працу з асобнымі гукавымі адзінкамі (фанемамі, дыфонамі) працягласцю 10–150 мс, на якім апрацоўваюцца спектральныя ахінальныя і каэфіцыенты MCEP, што забяспечваюць пераборлівасць гаворкі. Суперсегментны (прасадычны) узровень апісвае інтанацыйны контур выказвання ў цэлым – траекторыю $f_0(t)$, ахінальную інтэнсіўнасці $E(t)$, паўзы і працягласці складаў, ахопліваючы часавыя маштабы ад 50 мс (склад) да некалькіх секунд (фраза).

**Аднаўленне спектральнай ахінальнай і сінтэз сігналу ўзбуджэння.** Акустычная мадэль на кожны фрэйм працягласцю 5 мс выдае вектар мел-кепстральных каэфіцыентаў памернасцю $D = 80$. Пераўтварэнне гэтых каэфіцыентаў у спектральную ахінальную $|H(\omega)|$ выконваецца праз мел-частотны кепстр:

$$\log\left|H(\omega)\right| = \sum_{k=0}^{D-1} c_k \cdot \cos\left(k \cdot a(\omega)\right),$$

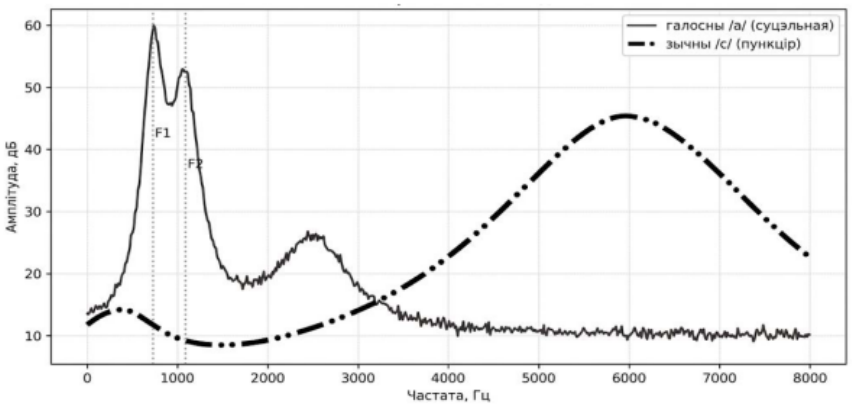
дзе $c_k$ – мел-кепстральныя каэфіцыенты; $a(\omega)$ – мел-адлюстраванне частаты:

$$a(\omega) = 2\arctan\frac{(a_\omega \sin\omega)}{(1 - a_\omega \cos\omega)} \approx 0{,}58 \text{ для } = 22050 \text{ Гц}.$$

Пасля пераўтварэння атрыманыя значэнні $\left|H(\omega)\right|$ нармуюцца з улікам мэтавай магутнасці фрэйма, якая задаецца прасадычным прэдыктарам праз параметр энергіі $E$:

$$H_{\text{norm}}(\omega) = H(\omega)\sqrt{\frac{E}{\sum_\omega \left|H(\omega)\right|^2}}.$$

Спектральная ахінальная для гукаў /a/ і /c/ дадзена на мал. 2.




*Мал. 2*. Спектральная ахінальная для гукаў /a/ і /c/

Сігнал узбуджэння $u(t)$ будуецца як лінейная суперпазіцыя перыядычнага (voiced) і шумавога (unvoiced) складнікаў з вагавымі каэфіцыентамі, якія задаюцца параметрам аперыядычнасці $BAP(t) \in [0, 1]$:

$$u(t) = \sqrt{1 - BAP(t)}\, u_p(t) + \sqrt{BAP(t)}\, u_n(t),$$

дзе $u_p(t)$ – перыядычны кампанент (імпульсны сігнал з перыядам $T = 1/f_0$); ), $u_n(t)$ – белы гаўсаў шум з адзінкавай дысперсіяй.

Перыядычны кампанент фарміруецца як паслядоўнасць сінхранізаваных глатальных імпульсаў формы Клата–Лільянкранца (Klatt-Liljencrants), а частата іх ідзе пад кіраваннем назапашвальнага лічыльніка фазы, што забяспечвае плыўнае змяненне $f_0$ без разрываў сігналу.

**Сінтэз фрэйма і прасадычная мадуляцыя.** Сінтэз кожнага фрэйма выконваецца метадам overlap-add (OLA) з акном Хана даўжыней $L = 512$ адлікаў і крокам $H = 110$ адлікаў (5 мс пры $f_s = 22050$ Гц), пасля чаго выніковыя фрэймы сумуюцца з нармаваннем, якое прадухіляе амплітудныя артэфакты пры перакрыцці вокнаў.

Контур асноўнага тону $f_0(t)$ з'яўляецца галоўным носьбітам інтанацыйнай інфармацыі. Прасадычны прэдыктар фармуе дыскрэтны контур $f_0[k]$ з крокам 5 мс, які затым лагарыфмічна інтэрпалюецца паміж ключавымі кропкамі (для захавання перцэпцыйнай лінейнасці шкалы паўтонаў) і згладжваецца фільтрам першага парадку з пастаяннай часу $\tau_f_0 \approx 8$ мс, што падаўляе высокачастотныя калыханні, уласцівыя дыскрэтнай акустычнай мадэлі.

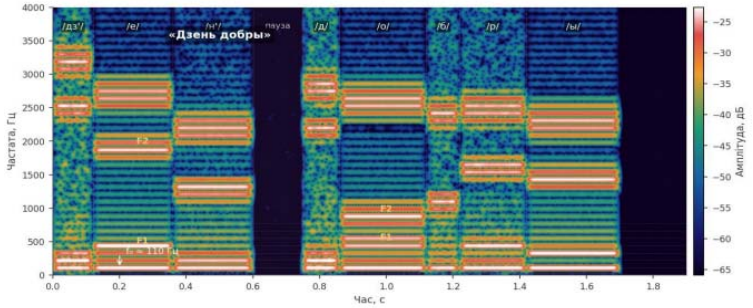
Ахінальная інтэнсіўнасці $E(t)$ адлюстроўвае ўдарна-рытмічную структуру фразы: націсканыя склады вылучаюцца падвышанай інтэнсіўнасцю, ненаціскныя – паніжанай. Значэнні $E[k]$, зададзеныя ў дэцыбелах, пераводзяцца ў лінейны маштаб і ўжываюцца да сігналу праз мультыплікатыўную мадуляцыю з косінусным кросфэйдам паміж суседнімі фрэймамі, што пазбягае клікаў на межах.

Кіраванне працягласцямі сегментаў рэалізуецца алгарытмам PSOLA (Pitch-Synchronous Overlap-Add): пры павелічэнні працягласці сегмента фрэймы дублююцца, пры памяншэнні – прапускаюцца, прычым маркеры фрэймаў прывязваюцца да найбліжэйшых глатальных імпульсаў, што захоўвае пітч-сінхроннасць і выключае перцэпцыйна прыкметныя артэфакты на стыках. Паўзы ўстаўляюцца ў выхадны буфер як блокі нулявых (сінтэтычныя паўзы) або нізкаамплітудных шумавых (запоўненыя паўзы) адлікаў зададзенай працягласці на аснове сінтаксічнага аналізу тэксту.

**Мікшыраванне сегментнага і суперсегментнага патокаў.** Пад мікшыраваннем разумеецца аб'яднанне двух незалежна сфармаваных патокаў даных – сегментнага сінтэзаванага сігналу і суперсегментных кіравальных параметраў – у адзіны PCM-струмень. Алгарытм працуе ў рэальным часе на ўзроўні буфераў гукавой карты (512–2048 адлікаў, што адпавядае 23–93 мс пры $f_s = 22050$ Гц) і прымяняе суперсегментныя параметры да сегментнага сігналу ў строгай паслядоўнасці: кіраванне перыядычнасцю ўзбуджальніка, кіраванне спектральнай ахінальнай праз MCEP бягучай фанемы, мадуляцыя аперыядычнасці, маштабаванне па працягласці (PSOLA), мультыплікатыўная мадуляцыя інтэнсіўнасці і, нарэшце, устаўка паўз.

Ключавой задачай пры гэтым з'яўляецца дакладнае часавае выраўноўванне прасадычных падзей (пікаў $f_0$, нарастанняў інтэнсіўнасці) з адпаведнымі націскнымі складамі, паколькі рассінхранізацыя больш за ± 20 мс перцэпцыйна прыкметная слухачу. Для гэтага для кожнай фанемы вызначаюцца адзнакі часу пачатку і канца ў адзінках фрэймаў вакодэра, а прасадычныя параметры індэксуюцца ў адзінках рэальнага часу, а не фанем ці фрэймаў, што сінхранізуе абодва патокі праз агульную часавую шкалу. Пры пераходзе паміж суседнімі фанемамі прымяняецца спектральны кросфэйд даўжыней 10 мс, які прадухіляе рэзкія скачкі спектральнай ахінальнай (мал. 3).

Перадача гатовых PCM-даных на гукавую карту рэалізавана праз падвойную буферызацыю з кальцавым буферам емістасцю ў чатыры блокі; сінхранізацыя паміж струменем сінтэзу і аўдыекалбекам, які выконваецца з рэальначасавым прыярытэтам, ажыццяўляецца атамарнымі індэксамі чытання і запісу. Для Windows выкарыстоўваецца WASAPI Exclusive Mode, для Linux – ALSA праз snd_pcm_writei(). Разлічаная затрымка сістэмы складае каля 11,6 мс, што значна ніжэй перцэпцыйнага парогу ў ~20 мс і забяспечвае прыдатнасць рашэння для інтэрактыўных дадаткаў.




*Мал. 3*. Спектраграма сінтэзаванай фразы з бачнымі фармантнымі пераходамі на межах фанем

Такім чынам, прапанаваная рэалізацыя вакодэра забяспечвае аднаўленне якаснага маўленчага сігналу з натуральнай інтанацыяй за кошт выразнага падзелу сегментнай і суперсегментнай апрацоўкі і іх строга ўпарадкаванага мікшыравання. Выкарыстанне алгарытмаў OLA, PSOLA і спектральнага кросфэйда дазваляе пазбегнуць пераходных артэфактаў, а падвойная буферызацыя забяспечвае працу сістэмы ў рэальным часе з непрыкметнай для слыху затрымкай.

Літаратура

1. Morise, M. WORLD: a vocoder-based high-quality speech synthesis system for real-time applications / M. Morise, F. Yokomori, K. Ozawa // IEICE Transactions on Information and Systems. – 2016. – Vol. E99-D, N 7. – P. 1877–1884.
2. Moulines, E. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones / E. Moulines, F. Charpentier // Speech Communication. – 1990. – Vol. 9, N 5–6. – P. 453–467.
3. FastSpeech 2: Fast and High-Quality End-to-End Text to Speech / Y. Ren [et al.] // Proceedings of ICLR. – 2021.
4. Klatt, D. H. Software for a cascade/parallel formant synthesizer / D. H. Klatt // Journal of the Acoustical Society of America. – 1980. – Vol. 67, N 3. – P. 971–995.
5. Рабинер, Л. Теория и применение цифровой обработки сигналов : пер. с англ. / Л. Рабинер, Б. Голд ; под ред. Ю. Н. Александрова. – М. : Мир, 1978. – 848 с.
6. Айфичер, Э. Цифровая обработка сигналов: практический подход / Э. Айфичер, Б. Джервис. – 2-е изд. – М. : Вильямс, 2004. – 992 с.

## МАКСИСЕЛЕКТОР ПИТАНИЯ ОТ СТАЦИОНАРНОЙ ЭЛЕКТРОСЕТИ И ФОТОЭЛЕКТРИЧЕСКОГО ПРЕОБРАЗОВАТЕЛЯ

**В. С. Тимофеенко, Н. В. Туровец**

*Учреждение образования «Гомельский государственный технический университет имени П. О. Сухого», Республика Беларусь*

Научный руководитель А. Е. Запольский

*Рассмотрена разработка интеллектуального максиселектора питания, который обеспечивает автоматическое переключение нагрузки между стационарной электросетью и фотоэлектрическими преобразователями в составе гибридной системы электроснабжения. Приведены принципы работы ключевых компонентов (фотопреобразователей, инверторов, контроллеров заряда), алгоритм управления на основе анализа напряжения в реальном времени, а также дана оценка экологических показателей.*

**Ключевые слова:** максиселектор, гибридная система электроснабжения, фотоэлектрические преобразователи, инвертор, контроллер заряда, возобновляемая энергетика, энергоэффективность.