



## VIZUAL FONEMALAR VA SINXRONIZATSIYA USULI OG’IZ HARAKATLARI

**Beknazarova Saida Safibullayevna**

*Televizion texnologiyalari fakulteti “Audiovizual texnologiyalari” kafedrasini mudiri, t.f.d., professor*

Tel: +998903276666 [saida.beknazarova@tuit.uz](mailto:saida.beknazarova@tuit.uz)

**Kucharova Shaxlo Sobir qizi**

*“Audiovizual texnologiyalari” kafedrasini magistranti*

Tel: +998936897632 [shaxlosobirovna06@gmail.com](mailto:shaxlosobirovna06@gmail.com)

**Annotasiya:** Yaratilayotgan personaj animatsiyasida og’iz orqali takrorlanadigan tovushlar bor, ular uchun og’izning shakli qanday bo’lishidan qat’iy nazar, vizual fonemalar yordamida mutlaqo qayta yaratilishi mumkin. U tovushini talaffuz qilishda trubali (uzaytilgan) lablar, qisqa m tovushini talaffuz qilganda yopiq lablar harakati tasvirlanadi. Vizual fonemalar og’izni ochish, yopish va cho’zish, qisqarishga qaraganda ancha xilma-xildir. Ushbu maqolada vizual fonemalar yordamida sinxronlashtirilgan lablar yoritiladi.

**Kalit so’zlar:** vizual fonemalar, lablar, tovushlar, sinxronlash.

Inson nutqining tabiati bimodaldir [1]. Biror animatsiya personaji tomonidan kuzatilgan nutq unga bog’liq audio xususiyatlari, shuningdek, labda sinxronizatsiya yoki yuz kabi vizual xususiyatlarda ifodalarga bog’liq. Nutqning vizual xususiyatlari akustikadagi shovqinli muhit tufayli nutqning xususiyatlari yo’qotishlarni qoplashi mumkin. Bu eshitish va nutqni aniqlash faqat audio yoki faqat vizual xususiyatlardan ko’ra aniqroqdir. Idrok nutqni audio va kabi bir nechta axborot manbalaridan foydalanish orqali yaxshilash mumkin nutqning video xususiyatlari. Oxirgi yigirma yil ichida foydalanish bilan ko’plab tadqiqotlar olib borildi Nutqning bimodal tabiati real vaqt nutqiga asoslangan yuz animatsiyalari.

*1-jadval*

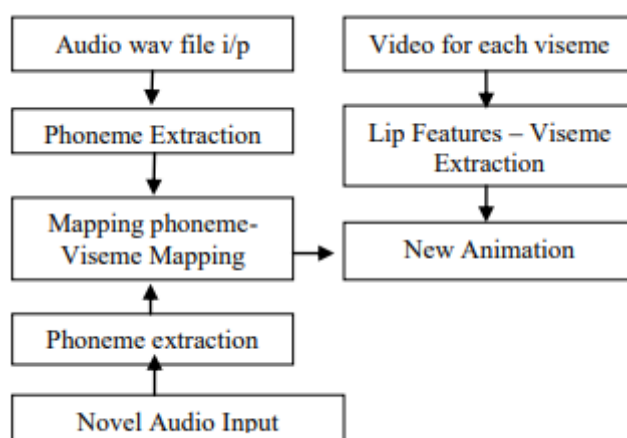
### *Vizual fonemalar*

№	Vizual fonemalar	Ovozli misollar	Qoidalar
1	b, m, p	bambuk, market, park	Lablar yopiq holda
2	i	video	Lablar cho’zilgan holda
3	f, v	fayl, avtomatik	Pastki lab bukilgan
4	u	uzoq	Lablar toraygan holda
5	r	rang	Boshqa ko’rinishlarga nisbatan ba’zan torroq, ba’zan kengroq
6	t, s	televizor, astronomiya	Boshqa ko’rinishlarga nisbatan ba’zan torroq, ba’zan kengroq

Lablar sinxronizatsiyasi - bu og’iz va tilning harakatini nutqda aniqlash [3]. Nutq tovushi tovush paychalarining tebranishi natijasida ovozli tovushlar va shivirlangan tovushlar holatida havo turbulenti hosil bo’ladi [4]. Ovoz ishlab chiqarilgan tomoq, til,

og'iz, tishlar, lablar va boshqalarni o'z ichiga olgan ovoz burun bo'shlig'i yo'lga asoslangan holda modellashtirilgan. Unli tovushlar nafasning nisbatan erkin o'tishi orqali hosil bo'ladi halqum va og'iz bo'shlig'i, undosh tovushlar esa qisman yoki to'liq hosil bo'ladi nutq organlarining har xil siqilishlari bilan havo oqimini to'sib qo'yadi.

Intonatsiya xarakteristikallari balandlik, amplituda va ovozi-pichirlash sifati va ular tovush manbaiga bog'liq bo'lsa, ovoz yo'li fonemani aniqlaydi. A fonema akustik nutqning asosiy birligidir. Fonemalarning vizual ifodalanishi viseme deb ataladi. Vizual jihatdan noaniq bo'lgan ko'plab akustik tovushlar mavjud. Shuning uchun, fonemalar va vizemalar o'rtasida ko'p-bir xaritalash mavjud. Dudak qilish uchun sinxronizatsiya mumkin, og'iz va tilning holati bilan bog'liq bo'lishi kerak nutq tovushining xususiyatlari (labda sinxronizatsiyaning asosiy g'oyasi ko'rsatilgan 1-rasm. Holbuki, og'iz va tilning pozitsiyalari fonemaning vazifalari va nutq tovushining intonatsion xususiyatlaridan mustaqildir.



1-rasm. Lablarni sinxronlashtirishning asosiy g'oyasi.

Inson ovozini yuzga aylantirish uchun turli usullar taklif qilingan harakat va keyin nutq signallaridan yuz animatsiyasini boshqarish uchun [3-10, va 12]. Asosiy masala - akustik nutqni og'iz shakliga aylantiradigan audio-vizual xaritalash parametrlari.

Xususiyatlarni ajratib olish usuli xususiyatni ajratib olishning ikki darajasini o'z ichiga oladi. Birinchidan, fonema xususiyatlar yozib olingan audiodan olinadi va undan keyin vizual xususiyatlarni ajratib olish kerak Har bir fonema yoki fonema guruhi uchun yozib olingan video.

Audio funksiyalarni chiqarish usullari. Birinchi qadam fonemalarni ajratib olishdir. Fonemalar eng yaxshi til birliklari sifatida tavsiflanadi. Ular bizning so'zlarimizni hosil qilish uchun birlashadigan tovushlardir, ammo qanday qilib a Fonemaning tovushga aylanishi ko'pgina omillarga, jumladan, atrofdagilarga bog'liq fonemalar, so'zlovchi urg'usi va yoshi. Ingliz tili yaxshilik qilish uchun 500 000 yoki undan ortiq so'zlarni etkazish uchun 40 ta fonemadan foydalanadi nutq mexanizmlari ishlaydigan ma'lumotlar elementi.

Nutqni ajratib olishda unli tovushlar asosiy yordam beruvchilar hisoblanadi. Ular quyidagicha ham ifodalanadi, og'iz ochiq tovushlar: a, o, ou.

2-jadvalda vizual fonemalardan og'izni cho'zish, toraytirish, ochish, yopishning individual shakllari o'rtasidagi ba'zi farqlarni ko'rsatiladi. Bu farqlarga qaramay, biz hali

ham qila olamiz qandaydir tarzda bunday tasvirlash shakllari yordamida vizual fonemalarni ifodalaydi.

*Vizual fonemalarni cho'zilgan shakllar bilan ifodalash, siqilish, og'izni ochish-yopish*

№	Vizualfonemalar	Ta'rifi	Sxematik tasvir
1	b, m, p	Lablar yopiq	
2	i / cho'zish	Og'iz ochiq va lablar cho'zilgan	
3	f, v	Pastki lab bukilgan	
4	u	Og'iz ochiq lablar biroz toraygan	
5	i	Og'iz ochiq lablar biroz cho'zilgan	
6	r	Boshqa ko'rinishlarga nisbatan ba'zan torroq, agar toraymagan bo'lsa	
7	t, s	Boshqa ko'rinishlarga nisbatan ba'zan kengroq, agar cho'zilmagan bo'lsa	

Akustik nutq va tegishli vizual nutq harakatlari o'rtasidagi xaritalash neyron tarmog'ini o'rgatish orqali aniqlash mumkin [10]. Trening bosqichida, kiritish naqsh va chiqish naqshlari tarmoqqa taqdim etiladi. Tegishli texnologiya ham chunki yashirin qatlamlar soni va qatlamdagi tugunlar soni aniqlanishi kerak. Yagona tarmoq barcha vizual parametrlarni yoki ko'plab tarmoqlarni qayta ishlab chiqarishga o'rgatilgan bo'lishi mumkin. Har bir tarmoq bitta vizual parametrlarni baholaydigan qilib o'rgatish mumkin. Neyron tarmoqlarni audio va vizual xaritalash uchun o'rgatish mumkin, shunda ular hisobga olinadi audio kontekstli ma'lumotlar (masalan, vaqtni kechiktiruvchi neyron tarmoqlar - TDNN). TDNN ko'proq hisoblash jihatidan HMM ga qaraganda samarali, lekin ko'p sonli yashirin birliklarni talab qiladi, qaysi o'qitish bosqichida yuqori hisoblash murakkabligiga olib keladi. Ko'pgina yondashuvlar a dan turli xil texnikalarning kombinatsiyasi foydalanadi. Biz neyron tarmoqni xaritalash uchun fonema va vizema ishlatdik.

**ADABIYOTLAR:**

[1] T. Chen and R. Rao, "Audio-visual integration in multimodal communication", Proceedings of IEEE, Special Issue on Multimedia Signal Processing, (1998) May, pp. 837-852.

[2] P. Vanroose, G. A. Kalberer, P. Wambacq and L. Gool, "From speech to 3D face animation", Proceedings of the Benelux Symposium on Information Theory, (2002).

[3] D. F. McAllister, R. D. Rodman, D. L. Bitzer and A. S. Freeman, "Lip synchronization for Animation", Proceedings of SIGGRAPH 97, Los Angeles, CA, (1997).

[4] J. P. Lewis and F. I. Parke, "Automated lip-synch and speech synthesis for character animation", Proceedings of SIGGRAPH, (1990).



[5] S. Kshirsagar and N. Magnenat-Thalmann, "Lip synchronization using linear predictive analysis", Proceedings of IEEE International Conference on Multimedia and Expo, New York, (2000).

[6] F. J. Huang, T. Chen, "Real-time lip-synch face animation driven by human voice", Proceedings of IEEE Multimedia Signal Processing Workshop, Los Angeles, California, (1998).

[7] Y. Li, F. Yu, Y. Xu, E. Chang and H. Shum, "Speech-driven cartoon animation with emotions", Proceedings of the ninth ACM international conference on Multimedia, Ottawa, Canada, (2001).

[8] M. Brand, "Voice Puppetry", Proceedings of SIGGRAPH'99, (1999).

[9] Y. Huang, X. Ding, B. Guo and H. Shum, "Real-time face synthesis driven by voice", Proceedings of Computer-Aided Design and Computer Graphics, Kunming, PRC, (2001).

[10] D. W. Massaro, J. Beskow, M. M. Cohen, C. L. Fry and T. Rodriguez, "Picture my voice: Audio to visual speech synthesis using artificial neural networks", Proceedings of AVSP'99, Santa Cruz, California, (1999).

[11] I. S. Pandžić and R. Forchheimer, "MPEG-4 Facial Animation - The Standard, Implementation and Applications", John Wiley & Sons Ltd, (2002).

[12] P. Hong, Z. Wen and T. S. Huang, "Real-time speech driven avatar with constant short time delay", Proceedings of International Conference on Augmented, Virtual Environments and 3D Imaging, Greece, (2001).

[13] Ph.D. Thesis, "Animating Faces from Speech", By Gwenn Englebienne, Faculty of Engineering and Physical Sciences, (2008).