

# Монгол бичвэрээс яриа үүсгэх үүсгүүр ба түүний үнэлгээ

К.Есенбек, Ч.Алтангэрэл, Ж.Пүрэв

Монгол Улсын Их Сургууль, Мэдээллийн Технологийн Сургууль  
Компьютер, Мэдээллийн Технологийн Тэнхим  
[esenbek@num.edu.mn](mailto:esenbek@num.edu.mn)

*Хураангуй* — Энэхүү өгүүлэл нь өгөгдөлд суурилсан монгол бичвэрээс яриа үүсгэх үүсгүүр болон түүний чанарын үнэлгээний талаар өгүүлдэг. Энэ ажлын зорилго нь хараагүй хүмүүст зориулсан дэлгэц унших програмын тусламжтайгаар хэрэглэх монгол бичвэрээс яриа үүсгэх үүсгүүрийг хөгжүүлэх явдал юм. Бичвэрээс яриа үүсгэх зорилгод тулгуурлан одоогийн байдлаар хамгийн орчин үеийн арга болох Далд Марковын загвар (HMM-Hidden-Markov-Model)-т суурилсан ярианы синтез буюу HTS-ийг хэрэглэсэн ба эмэгтэй хоолойгоор уншуулж, монгол хэлний бүх авиаг багтаасан 6 цаг гаруйн бичлэг бүхий ярианы корпус бэлтгэсэн. Үүнийг ашиглан эмэгтэй хоолойтой монгол хэлний бичвэрээс яриа үүсгэх үүсгүүр гарган авсан юм.

*Түлхүүр үгс*—Ярианы корпус, Бичвэрийг яриа болгох үүсгүүр

## I. УДИРТГАЛ

Бичвэрээс яриа үүсгэх үүсгүүр нь өгөгдсөн үгсийг автоматаар яриа болгон хөрвүүлдэг.

Монгол хэлний бичвэрээс яриа үүсгэх үүсгүүр нь ялангуяа харааны бэрхшээлтэй иргэдэд нэн шаардлагатай байгаа юм. Өмнө хэд хэдэн монгол бичвэрээс яриа үүсгэх үүсгүүр хөгжүүлсэн байдаг бөгөөд эдгээрээс цөөн тооны үүсгүүр амжилттай хөгжүүлэгдсэн байдаг. Тухайлбал дифонд суурилсан яриа үүсгүүр гэх мэт [1, 2]. Гэхдээ сайн бичвэрээс яриа үүсгэх систем нь харааны бэрхшээлтэй хүмүүс дэлгэц унших програм ашиглан хэрэглэх ёстой ба яриа нь тод ойлгомжтой байхыг шаарддаг.

Энэхүү бичвэрээс яриа үүсгэх зорилгод зориулан одоогийн байдлаар хамгийн орчин үеийн арга болох Далд Марковын загвар (HMM-Hidden-Markov-Model)-т суурилсан ярианы синтез буюу HTS [7] -ийг хэрэглэсэн. Уламжлалт аргаас ялгаатай нь, HTS зорилгот ярианы хэсгийг дүрслэн үзүүлэхдээ, ярианы дохионы магадлалт загвар болох акустик загвар (AM)-ын багцыг үүсгэж, дараа нь зорилгот ярианы дохиог үүсгэхийн тулд холбогдох акустик загваруудыг залгаж нийлүүлдэг. Зорилгот ярианы дохиог үүсгэхийн тулд HTS нь акустик загваруудын цувааг ярианы дохионд хувиргахдаа долгионы хэлбэрийг дахин байгуулах аргыг хэрэглэдэг.

Энэхүү өгүүлэл нь Далд Марковын загварт суурилсан бичвэрээс яриа үүсгэх үүсгүүрийг хөгжүүлсэн ярианы

корпус, хэрэгслүүд болон түүний чанарын үнэлгээний талаар талаар танилцуулах юм.

## II. МОНГОЛ ХЭЛ

Монгол хэл нь Алтай хэлний бүлэгт багтдаг бөгөөд олон зуун жилийн түүхтэй, үгийн сангийн баялаг өв сантай хэл юм. Одоогоор нийт 6 сая гаруй хүн монгол хэлээр ярьж байна. Үүнд, Монгол улсын 2,7 сая хүн, БНХАУ-ын Өвөр монголын өөртөө засах орон (ӨМӨЗО), Шинжаан, Төвдийн ойролцоо Хөх нуур муж, ОХУ-ын Буриад, Халимаг, Афганистаны Хярат, Будакшан, Майман зэрэг багтана. Монгол улсад халх, буриад, ойрд, цахар зэрэг олон аман аялгуу байдгаас халх аялгуу дийлэнх буюу 90 хувийг эзэлдэг.

Монгол хэл нь Алтай хэлний нийтлэг шинжийн дагуу залгамал хэл. Бүтээврүүд нь ар араасаа залгагдан орж үгийг бүтээдэг. Иймээс үг бүтээх болон хувилгах олон тооны дагавар, нөхцөл байдаг. Ойролцоогоор монгол хэлэнд 40,000 – 50,000 үг байдаг. Гэхдээ эдгээрээс зөвхөн 10,000 орчим л язгуур үг, 300 орчим дагавар юм [3].

### A. Монгол цагаан толгой

Монгол хэл нь Монгол бичиг (уйгаржин монгол бичиг), Латин, Кирилл гэсэн 3 төрлийн цагаан толгойтой. Монгол бичгийг XIII зуунаас ашиглаж ирсэн бөгөөд энэ нь хэсэг хугацаанд латин цагаан толгойгоор солигдон хэрэглэгдэж байжээ. 1940 оны засгийн газрын шийдвэрээр кирилл цагаан толгойг танилцуулж улмаар хуучин монгол бичгийг орлож Монгол улсын албан ёсны хэл болсон юм. Одоо Кирилл цагаан толгойг өдөр тутмын амьдрал болон компьютерийн системүүдэд ашиглаж байна[4,5].

### B. Монгол хэлний авиа зүй

Монгол хэлний үгийн үе нь дараах бүтэцтэй байна. Үүнд: V, CV, CVC, VVC, CVCC, VCCC, CVCCC; Энд “C” нь гийгүүлэгч ба “V” нь эгшиг юм. Монгол хэл нь 7 эгшиг, 7 урт эгшиг, 5 хос эгшиг, 5 туслах эгшигтэй. Монгол хэлний авиа зүйн хэллэг болон хувирлаас шалтгаалан 19-23 гийгүүлэгчтэй гэж үздэг [4]. Гийгүүлэгийг эр, эм гэж хоёр ангилна. Өөр нэг авиа зүйн онцлог нь өргөлт юм. Үгийн эхний үед орсон эгшиг болон урт эгшиг өргөлттэй дуудагддаг [3]. Үүнээс гадна монгол хэлний үгийн санд гадны төвд, санскрит, хятад, орос, эртний турк, англи зэрэг хэлнээс зээлдсэн авиа нилээд бий.

III. Монгол хэлний хөмрөг

A. Монгол хэлний үгийн сан

МУИС-ийн Компьютер хэл шинжлэлийн судалгааны төв нь нийтлэл болон сонин хэвлэл, уран зохиол, хуулиас цуглуулсан 5 сая үгтэй корпус үүсгэсэн байдаг.

B. Монгол хэлний ярианы сан

Олон зорилгоор цуглуулсан монгол хэлний ярианы сан байдаг. Гэвч эдгээрийг ярианы үүсгүүрт зориулаагүй юм. Тухайлбал: автомат яриа танигчид зориулан загварчилсан ярианы сан байдаг. Гэхдээ үүнийг яриа танигчид зориулсан учир өөр өөр уншигчаар уншуулсан ярианы өгөгдлүүдийг цуглуулсан байдаг. 10 үгийг 10 удаа 10 өөр уншигчаар уншуулсан. Нийт 80 уншигчийн уншсан ярианы өгөгдөл нь 60,000 wav файлаас бүрддэг юм[6].

IV. ЯРИАНЫ КОРПУСЫН ЗАГВАР

Дээр дурдсанаар Монгол хэлний яриа үүсгүүрт зориулсан том ярианы сан байдаггүй юм.

A. Монгол хэлний авиан зүйн галиглалт

Энэхүү ярианы сангийн загварыг Монгол хэлний албан ёсны аялгуу болох Халх аялгуун дээр тулгуурлан хийсэн. Хүснэгт 1 болон Хүснэгт 2-т монгол хэлний бүх гийгүүлэгч, эгшиг авианууд болон зээлдсэн авиануудыг харуулсан. Хүснэгт 1-т харуулсан жижиг үсгээр бичсэн авианууд сул дуудагдах бөгөөд том үсгээр бичсэн авианууд нь чангаар дуудагдана. Хүснэгт 2-т тодорхойлох цэггүй “:” бичсэн нь богино эгшиг, тодорхойлох цэгтэй “:” урт эгшиг болон “y” –г залгаж бичсэнээр хос эгшигийг харуулсан.

Бичвэрийг шууд авиан галигт хувиргахын тулд тэмдэглэгээг авианд хувиргах (G2P) хувиргуурыг хөгжүүлсэн. G2P хувиргуур бичвэрүүдийг хэрхэн дуудагдахыг шинжлэн дуудлагыг авианы дараалал хэлбэрээр шууд харгалзуулан тавьдаг. G2P нь ердийн үг, тоон тэмдэгт, тэмдэг, товчлол гэх мэт бүх төрлийн бичвэрийг боловсруулж гаргах юм.

B. Бичгийн хөмрөг бэлдэх

Энэхүү яриа үүсгүүрийн акустик загварыг сургахад зориулсан текстийг 5 сая үгтэй корпусаас дараах шаардлагуудын дагуу сонгон авсан юм. Үүнд:

- Монгол хэлний бүх авиануудыг багтаасан байх

- Уншигчид хангалттай хэмжээний унших материал бэлдэх

Үүний үр дүнд монгол хэлний бүх авиаг багтаасан 1500 авиа тэнцвэртэй өгүүлбэрийн багцыг цуглуулсан юм. Зураг 1-т 1500 өгүүлбэрийн жишээг харуулав.

Хүснэгт 1. Монгол хэлний гийгүүлэгч авиа

Авиан галиг	Дуудлага	Авиан галиг	Дуудлага
b	Б	B	Бь
c	Ц	C	Гь
d	Д	D	Дь
g	Г	G	Га
k	К	K	Кь
l	Л	L	Ль
m	М	M	Мь
n	Н	N	Нь
p	П	P	Пь
r	Р	R	Рь
s	С	S	Ш
t	Т	T	Ть
v	В	V	Вь
x	Х	X	Хь
f	Ф		
h	Ч		
j	Я,Ё,Е,Ю		
ng	Нг		
w	Ж		
z	З		

Хүснэгт 2. Монгол хэлний эгшиг авиа

Авиан галиг	Дуудлага	Авиан галиг	Дуудлага
a	а	a:	аа
e	э	e:	ээ
i	и	i:	ий
o	о	o:	оо
u	у	u:	уу
@	ө	@:	өө
^	ү	^:	үү
y	ы	y	ы
ay	ай		
oy	ой		
uy	уй		
ey	эй		
^y	үй		

```
@l-ng|g-@0-r|s-@0-ng| s-a|l|r-y0-ng| x-o|l|j-o0-r| t-e-l-g| d-u-l-ng|d-u:0-r| z-u|l|r-a:0-s| n-o:l-s| z-a-l-r|l-a0-g|d-a0-w| ...
@l-ng|g-@0-r|s-@0-ng| x-o|l|j-o0-r| t-e-l-g| t-e-l-g| d-@l|r-@0-v| o|l|n-y0| u-l-l|s-y0-ng| i-l-x| x-u|l|r-a0-l| d-u-l-ng|...
d-a|l|x-i0-a0-d| h| t-a-l-v| t-e-l-g| ^y|l-d|v-e0|r-i:0-ng| e-l-r|x-i:0-g| b-^l-r|m-@0|s-@0-ng| x-a:l-x| s-a|l|n-a0|l-y0-g| ...
...
...
```

Зураг 1. 1500 авиан тэнцвэртэй өгүүлбэрийн авиан галигт хувиргалтын жишээ

(NUM_TSEVELMAA_0001	"өнгөрсөн сарын [sp] хоёр тэг [sp] дундуур зураас [sp] ноос зарлагдаж [sp] ...")
(NUM_TSEVELMAA_0002	"өнгөрсөн хоёр тэг тэг [sp] дөрөв оны [sp] улсын их хурал [sp] ...")
(NUM_TSEVELMAA_0003	"дахиад ч [sp] тав тэг үйлдвэрийн эрхийг [sp] бүрмөсөн хаах саналыг [sp] ...")
	...

Зураг 2. Түр зогсолт оруулсан ярианы галиглалын жишээ

V. ЯРИАНЫ КОРПУС БАЙГУУЛАХ

A. Ярианы хэсэгчлэл

Монгол хэлний акустик загварыг сургах авиан хэсгүүдийг бэлдэхийн тулд авиан галиглалыг бичсэн бичлэгтэй зэрэгцүүлэх шаардлагатай юм. Үүнд НММ -д суурилсан хүчээр зэрэгцүүлэх акустик загварыг ашигласан. Энэхүү авианы зэрэгцүүлэлтийн нарийвчлалыг нэмэгдүүлэхийн тулд авиан галиглалд түр зогсолтыг гар аргаар нэмж өгсөн. Ойролцоогоор 14,581 түр зогсолтыг оруулж өгсөн ба Зураг 2 –д жишээг харуулав.

B. Яриа бичилт

Яриа бичилтийг монгол хэл шинжлэлийн мэргэжилтэн эмэгтэй уншигчаар дуу, шуугианаас тусгаарласан өрөөнд бичсэн.

Нийт 6 цагийн 1500 өгүүлбэрийн бичлэгийг 48 kHz давтамжтай, 16 бит/жишээ болон моно сувгаар, wave форматаар бичсэн. Бичлэг хийсэн ярианы өгөгдлийн яриа, шуугианы харьцаа нь ойролцоогоор 30-40 орчим dB.

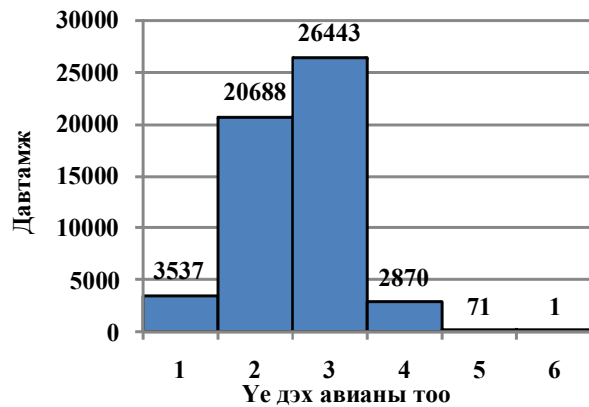
VI. КОРПУСЫН ШИНЖИЛГЭЭ

Үндсэн корпусын үзүүлэлтийг Хүснэгт 3-т харуулав. 6 цаг орчмын корпус нь ойролцоогоор 10,000 орчим давхардаагүй үгийг агуулсан байна. Зураг 3–т үеийн ургын гистограмыг харуулсан. Монгол хэлний авианы ургын гистограмыг Зураг 4–т харуулсан ба Хүснэгт 3-аас дундаж авианы урт нь 116,3 мс гэдгийг харж болно.

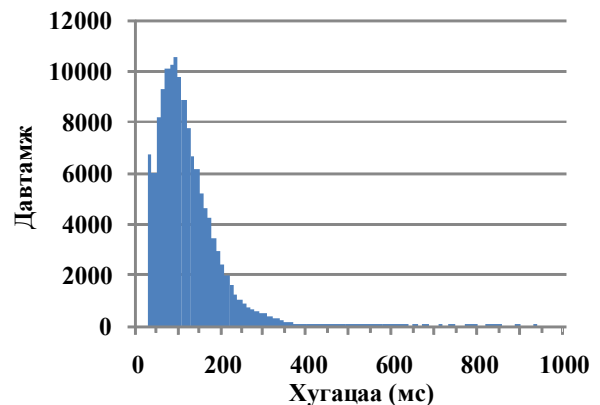
Зураг 5–д Монгол хэлний авианы тархацыг харуулав. Энэ авианы давтамжуудаас ажиглахад сул авианууд нь зүүн талд буюу их тархацтай байгааг харж болох юм. Харин эсрэгээр ихэнх чанга авианууд нь баруун талд буюу бага тархацтай байна. Эгшгүүдийн давтамжаас ажиглахад эгшгүүдийн тархац нь дараах дараалалтай байрласан байна. Үүнд: богино эгшгүүд (a, e, i, o, u), урт эгшгүүд (a:, e:, i:, o:, u:), ба хос эгшгүүд (au, eu, ou, eu, ^u).

Хүснэгт 3. Ярианы корпусын үзүүлэлт

Нийт урт	5 ц 52 мин
Өгүүлбэрийн дундаж урт	14.1 с
Хэлцийн дундаж урт	1.45 с
Үеийн дундаж урт	0.295 мс
Авианы дундаж урт	116.3 мс
Нийт өгүүлбэр	1,500
Нийт хэлц	14,581
Нийт үг	27,270
Давхардаагүй үг	9,401
Үе	53,610
Өргөлттэй үе	26,278
Өргөлтгүй үе	27,332
Нийт авиа	136,083
Үе дэх авианы дундаж	2.54
Давхардаагүй авиа	53

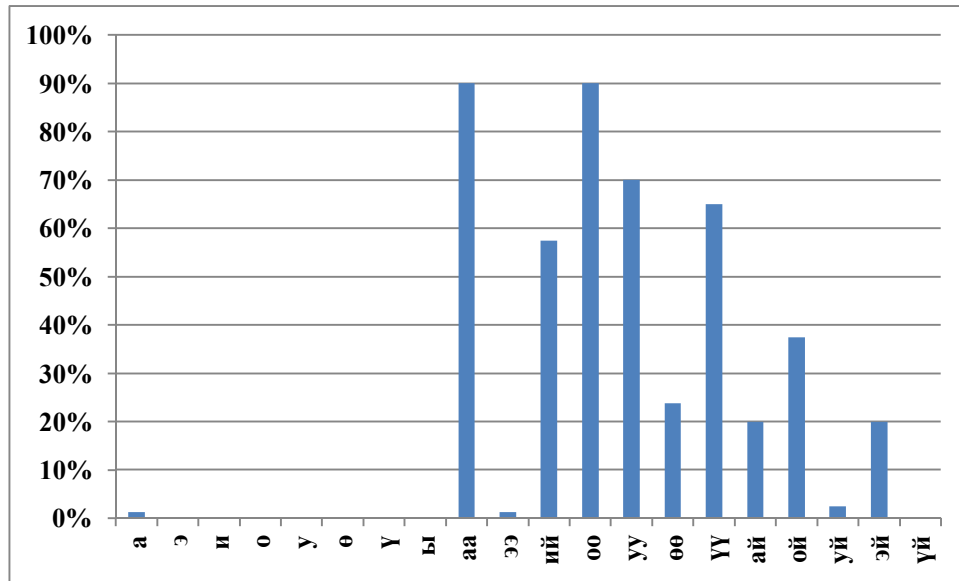


Зураг 3. Монгол хэлний үеийн гистограм



Зураг 4. Монгол хэлний авианы ургын гистограм





Зураг 8. Эгшиг авианы ойлгоц

Гийгүүлэгч, эгшиг авиануудын хувьд сонсоод ямар гийгүүлэгч, эгшиг авиа гэдгийг нь таних ёстой. Үр дүнг Зураг 7 –д гийгүүлэгч авиа, Зураг 8 –д эгшиг авиаг үзүүлэв. Эндээс гийгүүлэгч авианы хувьд Зураг 5-д авианы тархац нь бага, авианууд нь муу танигдсан гэж харагдаж байгаа юм. Эгшиг авиануудын хувьд урт, хос эгшгүүд танигдсан ба дангаараа дуудагдсан богино эгшгүүд муу танигдсан юм.

Өмнө хөгжүүлэгдсэн Дифонд суурилсан яриа үүсгүүрийн чанарын үнэлгээнээс харахад өгүүлбэрийн ойлгоц нь 60% орчим гарсан байдаг [2]. Мөн ярианы сонсогдох байдал, чанар, хурд зэргийг дээрхийн адил таван үзүүлэлтээр үнэлсэн байна. Эдгээр үзүүлэлтүүдтэй харьцуулахад бидний хөгжүүлсэн яриа үүсгүүрийн чанарын үнэлэмж нь өндөр байгаа юм.

VIII. Дүгнэлт

Энэхүү ажил нь өгөгдөлд суурилсан монгол бичвэрээс яриа үүсгэх үүсгүүр болон түүнд зориулсан ярианы хөмрөгийг хөгжүүлэх юм. Монгол хэлний бүх авиаг багтаасан ярианы хөмрөг үүсгэсэн бөгөөд үүнийг ашиглан ярианы үүсгүүрийг хөгжүүлэн ашиглаж байна. Гэхдээ үүнд шийдэх ёстой хоёр асуудал байгаа юм. Нэгдүгээрт G2P хувиргуурын гадны зээлсэн үгнүүдийг хувиргах асуудал, хоёрдугаарт акустик загварыг сургахад зарим авиануудын тоо цөөхөн мэт байгаа. Иймээс бид шинээр эргтэй болон эмэгтэй хоолойгоор уншуулсан, 3000 өгүүлбэр бүхий корпусыг бэлдэж байна.

ТАЛАРХАЛ

Энэхүү ажлыг хийхэд гүн туслалцаа үзүүлж G2P модулыг хөгжүүлсэн Ч.Тамир, өгүүлбэр уншсан Т.Цэвэлмаа нарт гүн талархал илэрхийлье.

НОМ ЗҮЙ

- [1] B. Batchuluun, D. Batchuluun, and, G. Gombosuren, “Mongolian text to speech conversion tool”, Project Ref.No: 0501A2\_L13, Interim report, InfoCon Co., Ltd, Mongolia, <http://www.infocon.mn/tts/>, 2006.
- [2] M. Davaatsagaan, and, K. Paliwal, “Diphone-Based Concatenative Speech Synthesis System for Mongolian”, Int’l MultiConference of Engineers and Computer Scientists (IMECS) 2008, Vol I, 2008.
- [3] D. Bayanduuren, “Mongolian Speech Recognition”, PAN localization Project Phase II, Mongolia, 2007.
- [4] A.Altangerel, and, B.Damdinsuren, “Research Report on ASR Development for Mongolian”, PAN localization Project Phase II, 2009.
- [5] Ц,Дамдинсүрэн, “Монгол шинэ үгсийн товч дүрэм”, Улаанбаатар, 1946.
- [6] A. Ayush, and, D. Bayanduren, “Mongolian Speech corpus”, PAN localization Project Phase II, 2009.
- [7] H.Chatchawarn, Th.Ausdang, Th.Nattanun, Ch.Altangerel, “Mongolian Speech Corpus for Text-To-Speech Development”, Speech Database and Assessments (Oriental COCOSDA), 2011 International Conference.