

Ярианы Дохионы Фрэйм Хоорондын LPC Болон Фурье Спектрийн Зайг Харьцуулсан Судалгаа

А.Мөнхбаясгалан, Д.Бямбажав

Электроник, Холбооны Инженерчлэлийн Тэнхим
Хэрэглээний Шинжлэх Ухаан, Инженерчлэлийн Сургууль
Монгол Улсын Их Сургууль
mbayasa_1014@yahoo.com

Хураангуй— Энэхүү судалгааны ажлаар бид ярианы дохионы фрэйм хоорондын Евклидийн зайг фурье спектр болон LPC спектруудийн хувьд харьцуулах ажлыг хийсэн. Энэхүү зайнууд нь нэг авианы дотор болон өөр өөр авианууд хооронд ямар байхыг тооцоолон харьцуулсан. Ингэж харьцуулахад зэргэлдээ фрэйм хоорондын LPC спектрийн зай нь уг болон эрэмбээс бараг хамаарахгүйгээр фурье спектрийн зайнаас 2.4 дахин бага харин өөр өөр авиа хоорондын LPC болон фурье спектрийн зайнууд нь ойролцоо гарсан. Энэ нь LPC спектрийн давуу талыг харуулж байна.

Турхуур үг— Ярианы дохио, фрэйм, LPC, фурье спектр, Евклидийн зай

I. УДИРТГАЛ

Өнөө үед ярианы дохиог ашиглаж хүн ба компьютерийн харилцааны ухаалаг системийг бий болгох судалгааны ажил өргөн хийгдэж хэрэглээнд нэвтэрч үр дүнгээ өгч байна. Ийм ухаалаг системийн нэг нь яриа таних систем юм.

1990-ээд оны сүүлчээс ярианы технологи өндөр хурдацтай хөгжсөний дунд хэрэглэгч өөрийн компьютертайгаа ярьж харин компьютер хүнд хариу хэлэх боломжтой болсон.

Ярианы технологи нь яриа таних систем, текстийг дуу авиа болгох систем, өгүүлэгчийг таних, өгүүлэгчийг батлах гэх мэт чиглэлүүдтэй [1]-[4]. Ярианы технологи нь хүн машин хооронд харилцах холбооны бүхий л хэлбэрт ашиглагдах боломжтой.

Энэ бүх технологиуд нь ярианы дохионд боловсруулалт хийж ярианы ялгагдах шинжийг (speech feature) гарган авах процессоос эхэлдэг.

Ялгагдах шинжийг гарган авах нь ярианы дохионы чадал, үелж буй импульсийн давтамж (pitch), формант болон өгүүлэхүйн эрхтэн системийн төлөв гэх мэт ялгаатай ялгагдах шинжүүдийг гарган авах процесс юм [5]. Ялгагдах шинжийг гарган авах нь олон үе шаттайгаар хийгддэг бөгөөд ярианы дохионд хамгийн эхний шатанд фурье анализ буюу шугаман таамаглалын коэфициентын (LPC) анализ хийх процесс хийгддэг [6].

Иймээс энэхүү ажлаар LPC-ийн аргыг сонгон ярианы дохионы фрэйм хоорондын зайг LPC болон фурье спектрийн хувьд харьцуулах ажлыг хийнэ.

II. LPC СПЕКТР

A. Шугаман таамаглал

Шугаман таамаглал буюу англиар Linear Prediction (LP) нь хугацааны хувьд тасралттай дохионы өмнөх түүврийн утгуудаар дараагийн утгуудыг шугаман функц ашиглан тооцоолох математик арга юм.

Ярианы дохионд анализ хийж яриаг чанар сайтай кодлох хамгийн ашигтай аргын нэг бол шугаман таамаглалын кодлол (LPC) бөгөөд ярианы параметруудийг оновчтой тооцоолж чаддаг. LPC-д Зураг 1-д үзүүлсэн загвараар яриа үүснэ гэж үздэг.

Өмнөх утгуудаар дараагийн утгуудыг тооцоолохдоо тэгшитгэл (1)-ийг ашигладаг.

$$s[n] = \sum_{k=1}^p a_k s[n-k] + e[n] \quad (1)$$

Үүнд: p - шугаман таамаглалын эрэмбэ, a_k - шугаман таамаглалын коэфициент, $e[n]$ нь таамаглалын алдаа юм.

Энэхүү загвар дахь шүүлтүүрийн дамжууллын функцыг

$$H(z) = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} = \frac{1}{A(z)} \quad (2)$$

гэж үздэг.

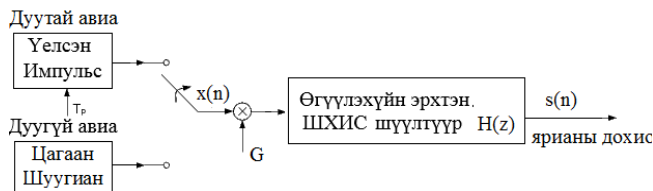
LPC –ийн аргыг яриа кодлох, ярианы синтез хийх, яриа таних, өгүүлэгчийг таних, өгүүлэгчийг батлах болон яриа хадгалахад өргөн хэрэглэж байна .

B. LPC спектр

Хүний яриа нь хүний хоолойд үүссэн өдөөх дохиогоор бий болдог бөгөөд хүний дууны, хамрын болон залгиурын эрхтэнүүдийн оролцоотойгоор хувирч өөрчлөгдөж байдаг.

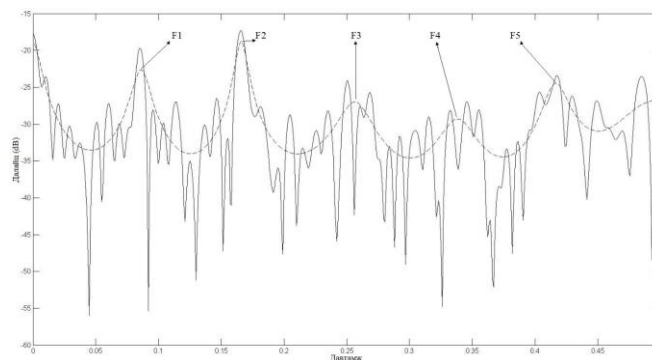
Хүний яриа нь өгүүлбэр, хэлц, үг, үеэс тогтох бөгөөд эдгээрийн суурь нь авиа юм. Зураг 1-д дуутай авиа (эгшиг авиа) хэлэгдэж байх буюу түлхүүр үелсэн дохионы үүсгүүр рүү залгагдсан байх үеийг харуулсан бөгөөд ямар эгшиг авиа хэлэгдэж байгаагаас хамаарч шүүлтүүрийн коэффициентүүд болох a_k нь өөрчлөгдөнө.

Зураг 1. Яриа үүсэх LP загвар



Амны болон залгиурын хэсгийн резонанс давтамжийг формант гэж нэрлэдэг. Ярианы дохионы хэсэгхэн хугацааны зурвас дахь буюу фрэймийн LPC спектрийг тасархай зураасаар, фурье спектрийг үргэлжилсэн зураасаар, формант давтамжуудыг (F1, F2, F3, F4, F5) Зураг 2-т харууллаа.

Зураг 2. Ярианы дохионы фрэймийн спектр (F1, F2, F3, F4, F5 нь формантууд)



III. Дохионы Фрэйм Хоорондын LPC Болон Фурье Спектрийн Зайг Харьцуулах

Энэхүү судалгааны ажлаар ярианы дохионы фрэйм хоорондын Евклидийн зайг LPC болон фурье спектрийн хувьд харьцуулсан.

A. Евклидийн зай

Хэрвээ \mathbf{p} болон \mathbf{q} нь m хэмжээст огторгуйн векторууд буюу

$\mathbf{p} = (p_1, p_2, \dots, p_m)$, $\mathbf{q} = (q_1, q_2, \dots, q_m)$ бол Евклидийн зай нь дараах тэгшитгэлээр бодогдоно.

$$d(\mathbf{p}, \mathbf{q})$$

$$= \sqrt{\sum_{i=1}^m (q_i - p_i)^2} \quad (3)$$

B. Фурье спектрийн Евклидийн зай

Ярианы дохиог фрэймүүдэд хувааж фрэйм бүрийн хувьд дискрет Фурье хувиргалтыг бодсон

бөгөөд нэг фрэймийн Фурье хувиргалтыг нэг вектор гэж үзсэн.

C. LPC спектрийн Евклидийн зай

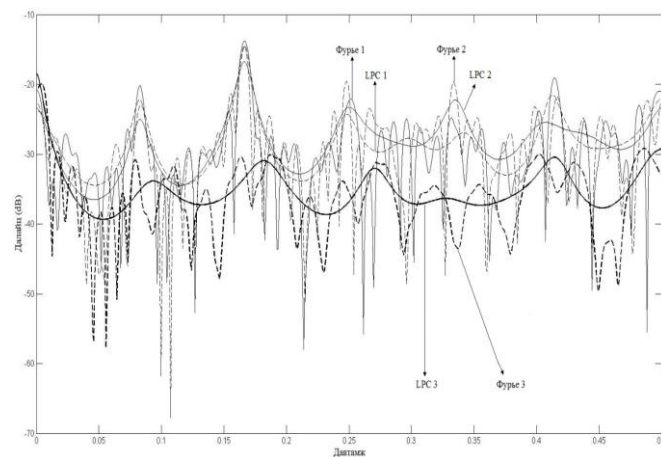
Мөн фрэйм бүрд LPC коэффициентүүдийг бодно. Тооцоолсон LPC коэффициентүүдээ ашиглан дамжууллын функцийг m тооны давтамж дээр бодож түүнийг нэг вектор гэж үзсэн.

IV. Туршилт Ба Үр Дүн

Фрэймийн уртаас хамаарч үүсэх фрэймийн тоо нь өөр өөр байна. 16 кГц-ээр түүвэрлэсэн ярианы дохионы LPC болон Фурье спектрийн хувьд Евклидийн зайг тооцоолон харьцуулан судалсан. Зураг 3-т жишээ болгон дараалсан хоёр фрэймийн фурье, LPC спектрууд болон хол байгаа фрэймүүдийн LPC болон Фурье спектрийг харьцуулан харууллаа. Зурагт тэмдэглэсэн LPC 1, болон LPC 2 нь дараалсан хоёр фрэймийн LPC спектр, LPC 3 нь хол байгаа фрэймүүдийн LPC спектр юм. Мөн адил Фурье спектрийг тэмдэглэсэн болно.

Хүснэгт 1-д нэг яриан дотор LPC спектруудийн Евклидийн зайг бүх зэргэлдээ фрэймүүдийн хувьд тооцоолоод тэдгээрийн дундаж утга болон стандарт хазайлтыг харууллаа. Үүнийг 6 үгний тохиолдолд LPC-ийн $p=16$, $p=20$, $p=32$ гэсэн 3 өөр эрэмбийн хувьд бодсон. Мөн Хүснэгт I-д Фурье спектрийн Евклидийн зайг бүх зэргэлдээ фрэймүүдийн хувьд тооцоолоод тэдгээрийн дундаж утга болон стандарт хазайлтыг тухайн үг бүрд харьцуулан харууллаа. Тодруулбал 19272 түүвэр урттай POD гэсэн үгийг 32 миллисекундээр 512 түүвэр урттай фрэймүүдэд хуваасан.

Зураг 3. LPC болон Фурье спектр



Ингэж хуваахдаа дараалсан фрэймүүдийг 64 түүврийн алхамтайгаар нийт 295 фрэйм үүсгэсэн бөгөөд 294 зэргэлдээ фрэйм байна гэсэн үг юм. Энэ 294 зэргэлдээ фрэймүүдийн хувьд LPC-ийн шугаман таамаглалын эрэмбэ $p=16$ үед LPC спектрийн Евклидийн зайн дундаж утгуудыг бодсон ба эдгээр утгуудын дундаж нь 0.8 гарсан. Мөн 294 зэргэлдээ фрэймүүдийн хувьд Фурье спектрийн Евклидийн зайн дундаж утгуудыг бодоод эдгээр утгуудын дундаж нь 2.1, стандарт хазайлтыг бодоод дундаж нь 0.6 байна.

Хүснэгт II-д нэг авиа дотор бус өөр өөр авианы фрэймүүдийн спектрийн зайнуудыг дээрхийн адилаар бодсон.

ХҮСНЭГТ I. НЭГ АВИАН ДОТОРХ СПЕКТРИЙН ЗАЙН ДУНДАЖ

Авиа	Евклидийн зай / нэг авиа доторх /							
	LPC p=16		LPC p=20		LPC p=32		Фурье спектр	
	Дундаж утга	Стандарт хязайлт	Дундаж утга	Стандарт хязайлт	Дундаж утга	Стандарт хязайлт	Дундаж утга	Стандарт хязайлт
CURD [kH@dH]	1.7	0.9	1.6	0.9	1.8	0.8	3.8	1.2
FED [fe:dH]	1.6	0.7	1.4	0.5	1.7	0.6	3.6	0.9
POD [pH@udH]	0.8	0.3	0.8	0.3	0.9	0.3	2.1	0.6
BAD [bAdHH]	1.2	0.4	1.5	0.4	1.3	0.3	3.1	0.8
SAID [sEdH]	2.2	1.0	1.9	0.8	2.2	0.9	4.9	1.6
DEAD [dHeidH]	1.1	0.3	1.0	0.3	1.1	0.4	2.6	0.8
Дундаж	1.4	0.6	1.4	0.5	1.5	0.5	3.3	1.0

УТГА, СТАНДАРТ ХАЗАЙЛТ

ХҮСНЭГТ II. ӨӨР ӨӨР АВИА ХООРОНДЫН СПЕКТРИЙН ЗАЙН ДУНДАЖ УТГА, СТАНДАРТ ХАЗАЙЛТ

Авиа	Евклидийн зай /өөр өөр авианууд хооронд /							
	LPC p=16		LPC p=20		LPC p=32		Фурье спектр	
	Дундаж утга	Стандарт хязайлт	Дундаж утга	Стандарт хязайлт	Дундаж утга	Стандарт хязайлт	Дундаж утга	Стандарт хязайлт
CURD [kH@dH]	24.6	7.9	24.6	7.9	24.5	7.9	24.8	7.9
FED [fe:dH]	16.3	4.7	16.3	4.7	16.2	4.7	16.5	4.7
POD [pH@udH]	11.4	3.6	11.5	3.6	11.5	3.6	11.8	3.6
BAD [bAdHH]	9.4	2.5	9.4	2.5	9.4	2.5	10.6	2.5
SAID [sEdH]	22.4	6.9	22.3	6.9	22.3	6.9	22.6	6.9
DEAD [dHeidH]	17.1	5.4	17.1	5.4	17.1	5.4	17.2	5.4
Дундаж	16.9	5.2	16.9	5.2	16.8	5.2	17.3	5.2

V. ДҮГНЭЛТ

Туршилтын үр дүнгээс харахад зэргэлдээ фрэйм хоорондын спектрийн зай нь дээрх 6 үг болон эрэмбээс бараг хамаарахгүйгээр LPC спектрийн Евклидийн зайн дундаж утга нь 0.8-2.2 хооронд байгаа бөгөөд дунджаар 1.4 байна. Харин Фурье спектрийн хувьд 3.3 байгаа бөгөөд LPC спектр хоорондын Евклидийн зай нь 2.4 дахин бага байна.

Өөр өөр авиа хоорондын спектрийн зай нь Фурье спектрийн хувьд Евклидийн зайн дундаж утга нь 10.6-24.8 хооронд байсан бөгөөд дунджаар 17.3 байхад LPC спектрийнх 16.9 буюу өөр өөр авиа хоорондын спектрийн зай нь Фурье болон LPC спектрийн хувьд ойролцоо гарсан.

Эцэст нь дүгнэхэд ярианы дохионд авианы анализ хийхэд LPC спектрийг ашиглах нь Фурье спектрээс илүү давуутай болох нь харагдлаа.

НОМ ЗҮЙ

- [1] Dr. Joseph Picone, Fundamentals of speech recognition: A short course, 1996
- [2] M.A. Anusuya, S.K. Katti, Speech Recognition by Machine: A Review, IJCSIS, Vol. 6, No. 3, 2009
- [3] Alan W Black, Speech Synthesis for Educational technology
- [4] Alan W black, Heiga Zen, Keiichi Tokuda, Statistical Parametric Speech Synthesis, 2011
- [5] Manish P. Kesarkar, Preeti Rao, Feature Extraction for Speech Recognition, 2003
- [6] J. Makhoul, "Linear Prediction: A tutorial review," Proc of the IEEE, vol. 63, no. 4, pp. 561-580, 1975.