

Rüstəmov Samir Sabir oğlunun Azərbaycan Respublikasının Prezidenti yanında Elmin İnkişafı Fondunun 5-ci “Mobillik qrantı” çərçivəsində “Neyro qeyri-səlis yanaşmaların nitqi tanıma sistemlərində tətbiqi” mövzusunda ABŞ-ın Florida Texnologiya İnstitutunda təcrübəkeçmənin nəticələri haqqında

## H E S A B A T I

Nitqi tanımanın əsas hədəflərindən biri insan və kompüter arasında ünsiyyətin tam olaraq nitq vasitəsilə həyata keçirilməsidir. Burada başlıca əngəl nitqi tanıma sistemlərinin işə başlamasında ilk olaraq qeyri-nitqin (mausun və ya klaviaturanın düyməsinin basılmasının) istifadə edilməsidir. Amma bildiyimiz kimi insanlar bir-birilə nitqli ünsiyyət qurduqda heç bir əlavə qeyri-nitq vasitəsindən istifadə etmirlər. İnsan-kompüter arasında ünsiyyəti tam olaraq nitqlə aparan, daim aktiv olan nitqi tanıma sistemlərinə ciddi ehtiyac duyulur. Yəni, kompyuter (və ya qurğu) həmişə xüsusi açar sözləri tanımaq üçün aktiv halda olur və bu söz tanındıqdan sonra kompyuterdə (və ya qurğuda) digər proqramlar işə düşür. Bu tip nitqi tanıma sistemləri “sözlə aktivləşmə” nitqi tanıma sistemləri adlanır və son illər geniş tədqiq olunan məsələlərdən biridir. Bu sistemlərə qoyulan əsas tələblər minimum enerji sərfiyyatına və yüksək tanıma dəqiqliyinə malik olmasıdır.

Layihə çərçivəsində əsasən iki tip nitqi tanıma (NT) sistemi– 1) “Sözlə aktivləşmə” (Wake-Up-Word) nitqi tanıma sistemi və 2) audio-video fayllarda açar sözlərin axtarılması (“Keyword Spotting”) məsələləri tədqiq olunmuşdur.

Nitqi tanıma sistemlərinin əsas tərkib hissəsi öyrətmə və tanıma proseslərindən ibarətdir. Hər iki prosesin girişində nitq signalını xarakterizə edən əsas əlamətlər hesablanır. Bu mərhələnin effektivliyi sonrakı mərhələlərin gedişinə və nitqi tanımanın dəqiqliyinə təsir göstərən mühüm amillərdən biridir. Nitq signalının informasiya daşıyan əl əlamətlərini seçmək lazımdır ki, signalı “ən yaxşı” təsvir edə bilsin. Layihə çərçivəsində nitq signalının əsas əlamətləri kimi MFCC (Mel Frequency Cepstral Coefficients) və LPC (Linear Predictive Coding) əlamətləri hesablanmışdır. Bu əlamətlərin əsas ideyasını nitq signalının səs tellərinin hasil etdiyi həyəcanlanma (excitation) signalının ( $e(n)$ ) səs yolu süzgecində (vocal tract filter– $\theta(n)$ )-də formalaşması təşkil edir:

$$s(n) = e(n) * \theta(n)$$

Signalın zaman oblastında bürünməsi, onların spektrlərinin tezlik oblastında hasilinə bərabərdir:

$$S(\omega) = E(\omega) \cdot \theta(\omega).$$

Nitqi əsas xarakterizə edən səs yolu süzgecidir ki, onu da həyəcanlanma signalından ayırmaq lazımdır. Bunun üçün MFCC üsulunda spektr loqarifmləyir:

$$\log|S(\omega)| = \log|E(\omega) \cdot \theta(\omega)| = \log|E(\omega)| + \log|\theta(\omega)|.$$

Tezlik oblastında tezliyin “böyük qiymətlərinə”  $E(\omega)$ , “kiçik qiymətlərinə” isə  $\theta(\omega)$  uyğun gəlir. Loqarifmlənən spektrə tərs Furiye çevirməsi tətbiq etməklə MFCC kepsstral əmsallar tapılır.

Eyni qayda ilə LPC üsulunda siqnalın zaman oblastında bürünməsinə  $z$ -çevirməsini tətbiq edilir:

$$S(z) = E(z) \cdot \theta(z)$$

Nitqi əsas xarakterizə edən səs yolu süzgəcini həyəcanlanma siqnalından ayırmaq üçün nitq siqnalının hər bir bölgüsü özündən əvvəlki  $p$  sayda bölgünün xətti kombinasiyası şəklində ifadə edilir:

$$\widehat{s}(a; n) = \sum_{k=1}^p s(n-k) \cdot a_p(k)$$

burada,  $a_p = (a_p(1), a_p(2), \dots, a_p(p))$  naməlum LPC əmsallar vektorudur.  $p$  tərtibi 8-12 arasında tam qiymətlər ala bilər ( $p \in [8, 12]$ ).

Həqiqi və əvvəlcədən deyilmiş bölgüləri cəmləyərək aşağıdakı siqnalı alırıq:

$$e(a; n) = s(n) + \widehat{s}(a; n) = s(n) + \sum_{k=1}^p a_p(k) \cdot s(n-k) \quad (1)$$

Bu siqnalı  $z$ -çevirməni tətbiq edək:  $R(z) = S(z)A(z)$ .  $A(z) = 1 + \sum_{k=1}^p a_p(k)z^{-k}$  funksiyası səs yolu süzgəcini xarakterizə edən  $\theta(z)$  funksiyasının tərs qiymətinə bərabərdir:  $A(z) = \frac{1}{\theta(z)}$ .

Səs yolu süzgəcini xarakterizə edən  $\theta(z)$  funksiyasını tapmaq üçün ilkin növbədə  $a_p$  əmsallarını tapmaq lazımdır. Bunun üçün aşağıdakı funksiyanı minimallaşdırmaq lazımdır

$$\varepsilon_p(a) = \sum_{n=0}^{\infty} |e(a; n)|^2 \rightarrow \min \quad (2)$$

Minimum zəruri şərtlərdən istifadə edərək məsələni həll edək:

$$\begin{aligned} \frac{\partial \varepsilon_p(a)}{\partial a_p(k)} &= \frac{\partial}{\partial a_p(k)} \sum_{n=0}^{\infty} |e(a; n)|^2 = 2 \sum_{n=0}^{\infty} e(a; n) \frac{\partial}{\partial a_p(k)} e(a; n) = \\ &= 2 \sum_{n=0}^{\infty} e(a; n) \frac{\partial}{\partial a_p(k)} \left[ s(n) + \sum_{l=1}^p a_p(l) s(n-l) \right] = 2 \sum_{n=0}^{\infty} e(a; n) s(n-k) = 0, \quad k = 1, 2, \dots, p, \end{aligned}$$

$$\text{Onda} \quad \sum_{n=0}^{\infty} \left[ s(n) + \sum_{l=1}^p a_p(l) s(n-l) \right] s(n-k) = 0 \quad (3)$$

alırıq.

$r_x(k) = \sum_{n=0}^{\infty} s(n)s(n-k)$  işarələməsi edək. Nəticədə (3) tənliyini aşağıdakı şəkildə yazmaq

bilərik

$$r_x(k) + \sum_{l=1}^p a_p(l)r_x(l-k) = 0 \text{ və ya } \sum_{l=1}^p a_p(l)r_x(k-l) = -r_x(k), \quad k=1, \dots, p \quad (4)$$

(4) tənliyi Yul-Volker (Yule-Walker) tənliyi adlanır.

(1) ifadəsini (2) funksionalında nəzərə alaraq:

$$\begin{aligned} \varepsilon_p(a) &= \sum_{n=0}^{\infty} |e(a;n)|^2 = \sum_{n=0}^{\infty} e(a;n)e(a;n) = \sum_{n=0}^{\infty} e(a;n) \left[ s(n) + \sum_{k=1}^p a_p(k)s(n-k) \right] = \\ &= \sum_{n=0}^{\infty} e(a;n)s(n) + \sum_{k=1}^p a_p(k) \sum_{n=0}^{\infty} e(a;n)s(n-k). \end{aligned}$$

$\sum_{n=0}^{\infty} e(a;n)s(n-k) = 0$  olduğundan, (2) funksionalını aşağıdakı şəkildə yazmaq olar

$$\varepsilon_{p,\min}(a) = \varepsilon_p(a) = \sum_{n=0}^{\infty} e(a;n)s(n) = \sum_{n=0}^{\infty} \left[ s(n) + \sum_{k=1}^p a_p(k)s(n-k) \right] s(n) = r_x(0) + \sum_{k=1}^p a_p(k)r_x(k).$$

Bu funksiya minimum verən  $a_p(k)$  -lar aşağıdakı Levinson-Durbin algoritmi ilə tapılır.

1. a)  $a_0(0) = 1$       b)  $E_0 = r_x(0)$
2.  $j = 0, 1, \dots, p-1$  üçün aşağıdakı ifadələr hesablanır

$$a) \quad \gamma_j = r_x(j+1) + \sum_{i=1}^j a_j(i)r_x(j-i+1)$$

$$b) \quad \Gamma_{j+1} = -\gamma_j / E_j$$

$$c) \quad i = 1, 2, \dots, j$$

$$a_{j+1}(i) = a_j(i) + \Gamma_{j+1}a_j(j-i+1)$$

$$d) \quad a_{j+1}(j+1) = \Gamma_{j+1}$$

$$e) \quad E_{j+1} = E_j \left[ 1 - |\Gamma_{j+1}|^2 \right].$$

Qurduğumuz nitqi tanıma sistemində nitqin əlamətlərinin hesablanması üçün LPC və MFCC üsullarının birgə istifadəsi tətbiq olunur. Nitqin əlamətlərinin seçilməsi algoritmi aşağıdakı kimi təyin edilir.

1. Nitq signalının ilkin emalı (Pre-processing):

Nitq signalı əsasən 4kHs-ə qədər tezlik zolağında dəyişir. Yüksək tezlikli əlavə küydən nitqi təmizləmək üçün signal FİR (Finite Impulse Response) süzgəcindən keçirilir:

$$s_{pe}(n) = s_{in}(n) - \alpha \cdot s_{in}(n-1)$$

burada,  $\alpha$  – süzgəcin əmsalıdır (bir çox sistemlərdə onu 0,95 və ya 0,97 qiymətlərindən istifadə olunur).  $s_{in}(n)$  – ilkin rəqəmli nitq signalıdır.

2. *Nitqin uc nöqtələrinin tapılması.* Tələffüz olunan nitq signalının əvvəli və sonunun tapılması nitqi tanıma sisteminin dəqiqliyinə təsir göstərir. Nitq signalının əvvəli və sonunun tapılması üçün istifadə olunan üsullardan biri VAD (Voice activation detection) üsuludur.

Bunun üçün signal bloklara bölünür. Bizim işlədiyimiz sistemdə blokun ölçüsü  $L=100$  bölgü (sample) götürülmüşdür. Hər blokun enerjisi ( $E_s$ ), gücü ( $P_s$ ), işarəsinin orta dəyişmə miqdarı ( $Z_s$ ) hesablanır:

$$E_s(m) = \sum_{n=m-L+1}^m s_p^2(n), \quad P_s(m) = \frac{1}{L} \sum_{n=m-L+1}^m s_p^2(n), \quad Z_s(m) = \frac{1}{L} \sum_{n=m-L+1}^m \frac{|\text{sgn}(s_p(n)) - \text{sgn}(s_p(n-1))|}{2},$$

burada,

$$\text{sgn}(s_p(n)) = \begin{cases} 1, & s_p(n) \geq 0, \\ -1, & s_p(n) < 0. \end{cases}$$

Fərz olunur ki, signalın ilk və ya son 5 bloku ətraf mühitin səsidir. Sonra isə aşağıdakı funksiyanın 5 bloka görə orta qiyməti ( $\mu_w$ ) və dəyişməsi ( $\delta_w$ ) hesablanır:

$$W_s(m) = P_s(m) \cdot (1 - Z_s(m)) \cdot S_c.$$

Bir çox sistemlərdə  $S_c = 1000$  götürülür,

$$t_w = \mu_w + \alpha \delta_w, \quad \alpha = 0,2 \cdot \delta_w^{-0,4}.$$

Nitqi ətraf mühitin səmindən ayıran funksiya aşağıdakı kimi qurulur:

$$VAD(m) = \begin{cases} 1, & W_s(m) \geq t_w, \\ 0, & W_s(m) < t_w. \end{cases}$$

Bu funksiyanın köməyi ilə asanlıqla nitq signalının uc nöqtələrini təyin etmək olar.

3. *Freymləmə (Framing).* Tanıma prosesinin dəqiqliyini artırmaq üçün nitq signalını qismən üst-üstə düşən freymlərə bölürlər.

$$s_{frame}(n) = s_{pe}(n) \cdot w(n),$$

$$w(n) = \begin{cases} 1, & K \cdot r < n \leq K \cdot r + N, \quad r = 0, 1, 2, \dots, M-1, \\ 0, & \text{qalan hallarda,} \end{cases}$$

burada,  $M$  - freymlərin sayı,  $f_s$  - diskretləmə tezliyi,  $t_{frame}$  - freymin zaman ölçüsü,  $K$  - freym addımı,  $N$  - isə freymin uzunluğudur:

$$N = f_s \cdot t_{frame}.$$

Bizim işlədiyimiz sistemdə diskretləşmə tezliyi  $f_s = 16kHs$ , freymin uzunluğu  $N=400$ , freymin addımı  $K=160$  götürülmüşdür.

4. *Pəncərələmə (Windowing)*. Sıqnal diskretləşdikdən sonra diskret Furiye çevirməsini (DFÇ) tətbiq etməzdən əvvəl, spektrin yayılmasını (spectrum leakage) azaltmaq üçün sıqnal ağırlıq funksiyası (weighting function) və ya pəncərə (window) adlanan funksiya vurulur. Geniş yayılmış pəncərə funksiyalarından biri də Hamminq funksiyasıdır:

$$s_w(n) = \left\{ 0,54 - 0,46 \cos\left(\frac{2\pi(n-1)}{N-1}\right) \right\} s_{frame}(n), \quad 1 \leq n \leq N.$$

#### MFCC əlamətlərin hesablanması.

*Diskret Furiye çevirməsi*. Hamarlanmış freymlərə diskret Furiye çevirməsi (DFÇ) tətbiq etməklə onların spektrləri hesablanır:

$$bin_k = \left| \sum_{n=1}^N s_w(n) e^{-i(n-1)k \frac{2\pi}{N}} \right|, \quad k = 0, 1, 2, \dots, N-1.$$

*Mel süzgəcləmə (Mel filtering)*. Mel süzgəcindən istifadə etməklə nitq sıqnalında olan faydalı informasiyanı itirməməklə verilənlərinin sayını azaltmaq olar:

$$Mel(x) = 2595 \lg\left(1 + \frac{x}{700}\right), \quad x = 700 \cdot \left(10^{\frac{mel}{2595}} - 1\right),$$

$$f_{c_i} = Mel^{-1}\left\{ Mel\{f_{start}\} + \frac{Mel\{f_s/2\} - Mel\{f_{start}\}}{NF} i \right\}, \quad i = 1, 2, 3, \dots, NF - 1,$$

$$cbin_0 = round\left\{ \frac{f_{start}}{f_s} N \right\},$$

burada, *round* - ədədə yaxın tam hissəni bildirir,  $NF$  - süzgəcin kanalların sayıdır,  $f_{start} = 64Hs$ . Bizim işlədiyimiz sistemdə  $NF = 24$  götürülüb:

$$cbin_i = round\left\{ \frac{f_{c_i}}{f_s} N \right\}; \quad cbin_{24} = round\left\{ \frac{f_s/2}{f_s} N \right\} = \frac{N}{2},$$

$$fbank_k = \sum_{i=cbin_{k-1}}^{cbin_k} \frac{i - cbin_{k-1} + 1}{cbin_k - cbin_{k-1} + 1} bin_i + \sum_{i=cbin_k+1}^{cbin_{k+1}} \left( 1 - \frac{i - cbin_k}{cbin_{k+1} - cbin_k + 1} \right) bin_i, \quad k = 1, 2, \dots, NF - 1.$$

*Loqarifmləmə*. Kepstralların alınması üçün spektr loqarifmlərin:

$$f_i = \ln(fb_{ank}_i), \quad i = 1, 2, \dots, NF - 1.$$

*Kepstral əmsallar*. Sıqnalı loqarifmlədikdən sonra tərs Furiye çevirməsi tətbiq etməklə mel tezlikli kepral əmsallar (MFCC) alınır:

$$C_i = \sum_{j=1}^{NF-1} f_j \cdot \cos\left(\frac{\pi \cdot i}{NF-1}(j-0.5)\right), \quad i = 0, 1, \dots, 12,$$

0-cı kepral az informativ olduğundan, əlamət vektorundan çıxarılır. Qalan 12 kepral isə sonrakı mərhələdə kepral orta çıxma tətbiq edilərək əlamət vektoruna daxil edilir.

*Kepral orta çıxma (Cepstral Mean Subtraction (CMS)).* Nitq signalı yazılarkən bəzi kanal səs-küyünə məruz qalır ki, bu da kanal effekti (channel effect) hadisəsi adlanır. Sistemin öyrədilməsi və sonradan istifadəsi zamanı ayrı-ayrı kanal effektləri sistemin tanıma dəqiqliyini aşağı salır. Bunu aradan qaldırmaq üçün kepral orta çıxmadan (CMS) istifadə olunur:

$$mc_j(q) = C_j(q) - \frac{1}{M} \sum_{i=1}^M C_i(q), \quad q = 1, 2, \dots, 12$$

**LPC əlamətlərin hesablanması.** Hər freymi ifadə edən LPC əmsallar Levinson-Durbin alqoritminin köməyi ilə tapılır. Tapılmış LPC əmsalların köməyi ilə freymlərin kepralları hesablanır.

$$c(k) = -a_p(k) - \sum_{i=1}^{k-1} \left(1 - \frac{i}{k}\right) a_p(i) c(k-i), \quad k = 1, \dots, 12$$

Alınan 12 LPC keprala sonrakı mərhələdə kepral orta çıxma tətbiq edilərək əlamət vektoruna daxil edilir.

Nitqi tanıma sistemlərinin əsasını təşkil edən müxtəlif riyazi modellər mövcuddur ki, onlardan geniş istifadə olunanlardan biri süni neyron şəbəkə (NŞ) modelidir. Tədqiq olunan məsələdə bir neçə süni neyron şəbəkənin birgə istifadəsi təklif edilmişdir.

Çoxlaylı süni neyron şəbəkə ümumi halda tam əlaqəli olmayan qrafdır. Tutaq ki,  $L$  – NŞ-nin laylarının sayı,  $N_\ell$  –  $l$ -ci layın neyronlarının sayı,  $l = 1..L$ ;  $I_{ij}^-$  –  $(l-1)$ -ci laydakı neyronlarla  $l$ -ci laydakı  $j$ -cu neyronla əlaqəli neyronlar;  $w_{ij}^\ell$  ( $l-1$ )-layın  $i$ -ci neyronu ilə,  $l$ -ci layın  $j$ -cu neyronu arasındakı əlaqənin çəkisidir (sinapsıdır). Neyron şəbəkəni öyrətmə prosesi zamanı  $l$ -ci laydakı  $j$ -cu neyron 0-cı laydakı giriş signaldan ( $x_p \in X$  vektorundan)  $s_{j,p}^\ell$  vəziyyətini qəbul edir və  $y_{j,p}^\ell$  çıxışını yaradır. Uyğun olaraq bütün şəbəkə üzrə  $y_p \in Y$  vektoru təyin olunur.  $X$  və  $Y$  çoxluqları uyğun olaraq NŞ-nin giriş və çıxışı üçün mümkün vektorların qiymətlərini təyin edir.

NŞ-də hesablanan  $s_{j,p}^l$  və  $y_{j,p}^l$  vektorlarının riyazi ifadəsini aşağıdakı şəkildə təsvir etmək olar.

$$s_{j,p}^l = \sum_{i \in I_{ij}^-} w_{ij}^l \cdot y_{i,p}^{l-1} + \theta_j^l, \quad (5)$$

$$y_{j,p}^l = f(s_{j,p}^l), \quad j = 1, \dots, N_l, \quad l = 1, \dots, L, \quad (6)$$

$$y_{j,p}^0 = x_{j,p}, \quad j = 1, \dots, N_0, \quad (7)$$

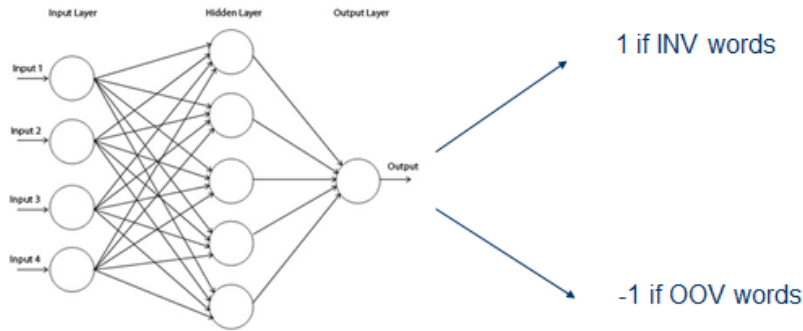
burada  $\theta$  –  $l$ -ci layın  $j$ -cu neyronunun kənara çıxmasıdır.  $f(\cdot)$  - verilmiş qeyri-xətti aktivləşmə funksiyasıdır. Aktivləşmə funksiyası kimi loqistik və ya hiperbolik tangens funksiyasını göstərmək olar.

Verilmiş məsələdə öyrədilən verilənləri  $\{x_p, d_p\}, p=1..P$  ilə işarə edək. Burada  $d_p = (d_{1,p}, \dots, d_{N_L,p})$  –  $x_p$  siqnalına uyğun NŞ-nin arzu edilən çıxış vektorudur. NŞ-nin öyrədilməsi ondan ibarətdir ki, elə  $w_{ij}^l$  çəkilmələri və  $\theta_j^l$  kənara çıxmalar  $i \in I_{ij}^l, j=1, \dots, N_l, l=1, \dots, L$  tapmaq lazımdır ki, NŞ-yə  $x_p$  giriş siqnalı verdikdə  $d_p$ -yə maksimal “yaxın” olan  $y_p$  çıxış siqnalı alınsın. Çox zaman öyrətmənin keyfiyyətini orta kvadratik səhv funksionalı ilə təyin edirlər:

$$E(w, \theta; x, s, y) = \frac{1}{P} \sum_{p=1}^P \eta_p E_p(w, \theta; x_p, s_p, y_p), \quad (8)$$

$$E_p(w, \theta; x_p, s_p, y_p) = \frac{1}{2} \sum_{j=1}^{N_L} (y_{j,p}^L - d_{j,p})^2,$$

burada  $\eta_p$  –  $p$ -ci surətin düzgün təyin olunmasının mümkünlüyünü təyin edən çəki əmsəlidir  $p=1, \dots, P, j=1, \dots, N_L$ . Surətlərin tanıma üçün qurulmuş neyron şəbəkənin öyrətmə məsələsi  $(w, \theta)$  parametrlə (8) kriteriyasını (5)-(7) asılılığı ilə minimallaşdırmaqdan ibarətdir. İşlətdiyimiz sistemdə neyron şəbəkələr qoşma qradiyent üsulu ilə öyrədilib. Tədqiq olunan məsələ iki sinfli klassifikasiya məsələsi olduğundan 1 çıxışdan ibarət neyron şəbəkə qurulmuşdur (şək 1). Yoxlanılmışdır ki, 1 çıxışlı neyron şəbəkə 2 çıxışlı neyron şəbəkədən daha yaxşı nəticə göstərir.



Şək. 1. NT sistemində istifadə olunan bir çıxışlı neyron şəbəkənin quruluşu

**Ekspərimentlərin nəticələri.** NT sistemin işini qiymətləndirilməsi üçün 2000 nitqli söz nümunəsi öyrətmə, 500 nümunə isə test prosesi üçün götürülmüşdür. Tanıma sistemi ayrı-ayrı bazalarda müxtəlif nəticələr göstərdiyindən test məsələni 10 dəfə ayrı-ayrı qruplaşdırılmış baza üçün yoxlanılmışdır (cədvəl 1-2). Precision - sistemin klassifikasiya etdikləri nümunələrin neçə faizinin doğru olduğunu, Recall isə doğru klassifikasiyanın neçə faizinin sisteminin təsdiqlədiyi göstəricidir.

Baza qrupları	Precision (%)	Recall (%)	F1 ölçü (%)	Dəqiqlik (%)
1	94,8	91,7	93,2	98,4
2	89,6	93,5	91,5	98,4
3	97,7	91,5	94,5	99
4	91,4	86,5	88,9	98,4

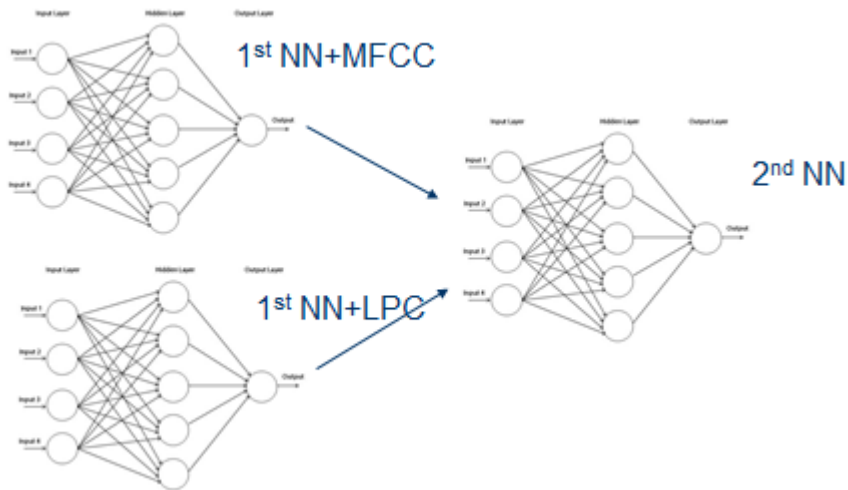
5	93,3	86,2	89,6	97,4
6	100	79,1	88,3	97,2
7	92	90,2	91,1	98,2
8	91,4	71,1	80	96,8
9	97,1	89,5	93,2	99
10	87,8	81,8	84,7	97,4
<b>Orta.</b>	<b>93,1</b>	<b>86,11</b>	<b>89,5</b>	<b>98,02</b>

Cədvəl 1. MFCC-əsaslı sözlə aktivləşmə NT sisteminin nəticələri

Baza qrupları	Precision (%)	Recall (%)	F1 ölçü (%)	Dəqiqlik (%)
1	80,8	70	75	94,4
2	95,6	93,5	94,5	99
3	81,6	85,1	83,3	96,8
4	88,2	80,1	84,5	97,8
5	83,3	69,2	75,6	94,2
6	96,3	77,6	86	96,6
7	85,1	78,4	81,6	96,4
8	96,9	68,9	80,5	97
9	93,5	76,3	84,1	97,8
10	95,1	88,6	91,8	98,6
<b>Orta.</b>	<b>89,64</b>	<b>78,77</b>	<b>83,69</b>	<b>96,86</b>

Cədvəl 2. LPC-əsaslı sözlə aktivləşmə NT sisteminin nəticələri

Cədvəl 1 və 2-də ifadə olunmuş nəticələrə görə yüksək dəqiqlik göstəricisinə malik olsa da 2-ci növ xətası yuxarıdır. 2-ci növ xətanı azaltmaq üçün öyrədici baza 2 hissəyə bölünmüşdür. 1000 nümunə öyrətmə, 1000 nümunə isə çarpaz qiymətləndirmə üçün götürülmüşdür. Öyrədilmiş bazaya görə çarpaz qiymətləndirmə bazası test edilmişdir və neyron şəbəkənin müxtəlif əlamətlərə görə nəticələri yeni qurulmuş ikinci neyron şəbəkənin girişinə verilmişdir (şək. 2). Əldə edilmiş nəticələr cədvəl 3-də verilmişdir.



Şəkil 2. Sözlə aktivləşmə sistemində istifadə olunan multi-strukturlu neyron şəbəkənin quruluşu

Baza qrupları	Precision (%)	Recall (%)	F1 ölçü (%)	Dəqiqlik (%)
1	80,8	98,3	88,7	97
2	74,6	95,7	83,8	96,6



3	78,6	93,6	85,4	97
4	76,1	94,6	84,3	97,4
5	77,9	92,3	84,5	95,6
6	91	91	91	97,6
7	78,8	80,4	79,6	95,8
8	61,9	86,7	72,2	94
9	62,3	86,8	72,5	95
10	86,7	88,6	87,6	97,8
<b>Orta.</b>	<b>76,87</b>	<b>90,8</b>	<b>82,96</b>	<b>96,38</b>

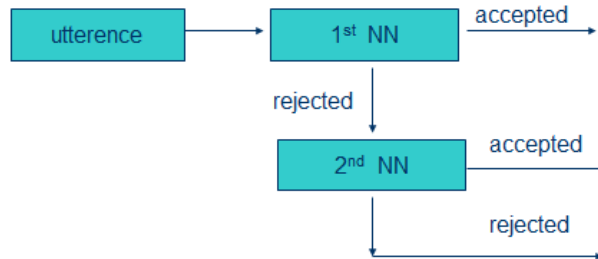
Cədvəl 3. Multi-strukturlu neyron şəbəkənin tanıma nəticələri.

Hər iki sistemin nəticələrini müqayisə etsək alırıq ki, birinci neyron şəbəkənin 1-ci növ xətası kiçik amma 2-ci növ xətası böyükdür. 2-ci neyron şəbəkədə isə bu göstəricilər əksdir (cədvəl 4).

Üsullar	Precision (%)	Recall (%)	F1 ölçü (%)	Dəqiqlik (%)
1-ci NŞ+MFCC	93,1	86,11	89,5	98,02
1-ci NŞ+LPC	89,64	78,77	83,69	96,86
2-ci NŞ	78,67	90,8	82,96	96,38

Cədvəl 4. 1-ci və 2-ci neyron şəbəkələrin tanıma nəticələri.

Tanıma dəqiqliyini artırmaq üçün bu iki sistem əsasında hibrid sistem qurulmuşdur (şək 3). Bu sistemdə tələffüz olunan nitq üçün ilk olaraq 1-ci neyron şəbəkə ilə tanıma həyata keçirilir (çünki onun 1-ci növ xətası kiçikdir). Bu şəbəkənin imtina etdiyi nümunəni isə 2-ci neyron şəbəkə yoxlayır. Bu halda həm birinci növ xətanı həm də ikinci növ xətanı azaltmış oluruq.



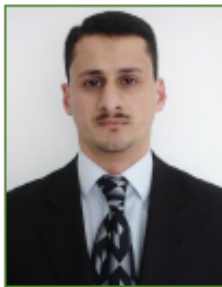
Şəkil 3. Sözlə aktivləşmə sistemin qərar qəbuletmə bloku.

**Qeyd.** Təcrübəkeçmə müddətində əldə edilmiş elmi nəticələr Florida Texnologiya İnstitutunun əsas elmi seminarında 13 fevral 2015-ci il tarixində məruzə edilmişdir.



## Florida Institute of Technology The Fredric M. Ham ECE Distinguished Seminar Series

### Wake-Up-Word Speech Recognition using Artificial Neural Network



Invited seminar talk by

**Dr. Samir Rustamov**

Associate Professor  
Institute of Control Systems  
Azerbaijan National Academy  
of Sciences



**Friday, February 13, 2015 @ 11:00–12:30**  
Room 273/276, Olin Engineering Complex, FIT, Melbourne Main Campus  
Hosted by Dr. Veton Kepuska, Associate Professor, ECE Dept.

#### Abstract

Wake-Up-Word (WUW) is a new paradigm in speech recognition (SR). WUW SR is defined as detection of a single word or phrase when spoken in the alerting context of requesting attention, while rejecting all other words, phrases, sounds, noises and other acoustic events and the same word or phrase spoken in non-alerting context with high accuracy. This talk is about investigation of WUW speech recognition. The study combines multiple speech features: MFCC and LPC for training and recognition. The system utilizes multi structured Artificial Neural Network. This work is done during the visit of Dr. Rustamov at FIT under the supervision of Dr. Këpuska.

#### Biographical Sketch

Samir Rustamov was born in Baku, Azerbaijan in 1980. He received B.S. degree in Applied Mathematics from Baku State University (BSU), in 2000, the M.S. degree in Optimization and Optimal Control from BSU, in 2002, and the Ph.D. degree in Computer Science from Institute of Cybernetics of Azerbaijan National Academy of Sciences (ANAS), in 2007. Between 2006 and 2012, he worked as scientific researcher at Institute of Cybernetics. In 2012-2013 he was post-doctoral researcher in School of ECE, Georgia Institute of Technology. At the present Dr. Rustamov is leading researcher at Institute of Control Systems of ANAS and lecturer at Qafqaz University. His research interests include Speech Recognition, Natural Language Processing and Machine Learning.

**Electrical & Computer Engineering**  
**Florida Institute of Technology**  
150 W. University Blvd  
Melbourne, Florida 32901  
(321) 674-8060  
<http://coe.fit.edu/ee/>

Please RSVP prior to attending this talk by visiting

[my.fit.edu/seces](http://my.fit.edu/seces)

Courtesy Parking Permits are available via our website.  
Finally, for latest SECES news and updates follow us on Twitter

@mySECES



# Wake-Up-Word Speech Recognition using Artificial Neural Network



**Samir Rustamov, PhD**  
Institute of Control Systems

13 February 2015  
Melbourne, Florida.

**Adviser: Prof. Veton Kepuska**  
Florida Institute of Technology

