

# Nitq texnologiyalarının eksperimental-fonetik aspektləri

Nəzakət Qaziyeva

AMEA Nəsimi adına Dilçilik İnstitutu, Bakı, Azərbaycan

*n.qazi@inbox.ru*

**Xülasə** — Məqalədə nitq texnologiyalarının inkişafında mühüm amillərdən olan eksperimental fonetik tədqiqatların əhəmiyyəti vurğulanır, nitq signalının akustik təhlilindən əldə olunmuş yeniliklərin bir çox müasir tədqiqat sahələri üçün vacib nəticələr verdiyi göstərilir. Təbii nitqə yaxın sintezin əldə olunması üçün nitq signalının fiziki-akustik parametrlərinin signalın mövqeyi, digər fonemlərlə əlaqəsi kimi amillər nəzərə alınmaqla əsaslandırılmasının verdiyi üstünlüklərdən bəhs olunur.

**Açar sözlər** — *nitq texnologiyaları; sintez; eksperimental fonetika; koartikulyasiya*

## I. GİRİŞ

Müasir dövrdə informasiya texnologiyalarının nailiyyətlərindən bəhrələnmək bir çox sahələr kimi, multidissiplinar istiqamət olan nitq texnologiyaları da sürətlə inkişaf etməkdədir. Dilin səs sistemi ilə əlaqəli biliklərin praktiki tətbiqi və insanın təbii nitq fəaliyyətini modelləşdirmə üsulları ilə bağlı olan nitq texnologiyalarının sürətli inkişafı müasir cəmiyyətin bu sahədəki bir sıra problemlərinin həllinə ehtiyacdən meydana çıxan tələblə əlaqədardır. Bu günə qədər dilçilik sahəsində təbii dillərin səs formasının təşkili və fəaliyyət prinsipləri haqqında xeyli məlumat toplanmışdır. Bu

məlumat nəzəri və eksperimental fonetik tədqiqatlar zamanı əldə edilmiş və əsasən təsviri xarakter daşmışdır. Müasir elmi baxışlar nitq texnologiyalarının fonetik məlumatlardan istifadə olunmaqla inkişafını aktuallaşdırır.

Nitq texnologiyalarına avtomatik nitqin tanınması və sintezi üçün sistemlərin inkişafı, nitq məlumatlarının korpusunun formalaşdırılması (akustik verilənlər bazası), nitq signalının avtomatik işlənməsi (əsas ton tezliyinin müəyyən edilməsi, formalı hərəkət trayektoriyalarının müəyyən edilməsi, avtomatik seqmentləşdirmə və s.), şəxsin səsle identifikasiyası və yoxlanılması, rabitə kanalları üzərindən ötürülməsi üçün nitq signalının sıxılması üsullarının işlənməsi və s. kimi məsələlər daxildir.

Nitq signalının rəqəmsal formada təqdim edilməsi onun təhlili və işlənməsi üçün geniş perspektivlər açır. Audio signalının səs analizi proqramları vasitəsilə təhlili zamanı səs signalının ossiloqramını və ya sonoqramını statik təsvir şəklində müşahidə etmək imkanı, signalı əvvəldən axıra və ya axırdan əvvələ doğru hərəkət etdirməyin mümkünlüyü, həmçinin həm bütün signalı, həm də onun hər hansı bir fraqmentini dəfələrlə dinləmək imkanı bir çox məsələlərin dəqiq təhlili və araşdırılmasına şərait yaradır. Bu seçilmiş fraqmenti, nəinki

dinləmək, həm də silmək, başqa yərə köçürmək və ya başqa səs signalına daxil etmək, həmçinin fiziki xüsusiyyətlərini dəyişdirmək mümkündür.

Eksperimental fonetikanın ənənəvi və aktual problemlərindən olan nitq axınının davamlılığı (qeyri-sabitliyi) ilə onun fonetik və fonoloji təsvirinin diskretliyi arasındakı ziddiyyət [1] bu məsələlərdə müəyyən problemlərin ortaya çıxmasına səbəb olsa da, nitq signalının akustik parametrlərinin eksperimental üsulla tədqiqinin müasir səviyyəsində bu tipli problemlərin həlli xüsusi çətinlik yaratmır. Bununla belə, müasir dövrdə bu sahə ilə bağlı yeni vəzifələr aktuallaşır. Bu problemləri həll etmək üçün artıq signalın spektral xüsusiyyətlərini dəqiq hesablamaq kifayət deyil, signalın qavrayış-perseptiv keyfiyyətini dəyişdirən amillər (əsas tezlik, formant mövqeyi və istiqaməti, fərdi tembr xüsusiyyətləri və s.) diqqət mərkəzində saxlanılmalıdır.

Nitq sintezi müəyyən informasiyanın çatdırılması tələb olunan istənilən sahədə tətbiq oluna bilər. Başlıca tətbiq sahələrindən isə informasiya-məlumat sistemləri, çağrı mərkəzləri, “Qaynar xətt” xidməti, sağlamlığı məhdud insanların, xüsusilə gözdən əlillərin cəmiyyətə inteqrasiyası məqsədilə “səsli” idarəetmə, nəqliyyat sektorunda – qatarların hərəkət qrafiki ilə bağlı, avtobus və metroda, eləcə də hava limanında müəyyən məlumatların verilməsi, eləcə də hərbi və aviakosmik sənaye qeyd oluna bilər.

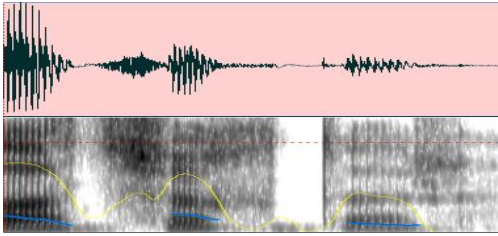
## II. “KLONLAŞDIRMA” METODUNUN TƏTBİQİNDƏ NİTQİN FİZİKİ PARAMETRLƏRİNİN ROLU

Son zamanlar nitqin sintezi üçün verilənlər bazasında saxlanılan nitq signalının seqmentlərinin birləşdirilməsindən ibarət olan konkatenativ üsuldan geniş istifadə olunur. Nitq məlumat bazasının həcmi nə qədər böyükdürsə, onda konkret şəxsə aid səs, temporal və intonasiya dəyişkənliyi ilə bağlı material bir o qədər tam əks olunur. Nəticədə, sintez edilmiş nitq daha asan “tanınır” və ümumiyyətlə daha təbii səslənir. Bundan əlavə, nitq korpusu nə qədər böyükdürsə, onun elementlərinin əsas tonunun (ƏTT) müddətini və tezliyini dəyişdirmək lazım olduqda baş verən nitq seqmentlərinin təhrifi bir o qədər az olur. Başqa sözlə, nitq verilənlər bazası nə qədər böyükdürsə, onun tələb olunan kontekstdə, tələb olunan müddət və tezlik konturlu elementi ehtiva etmə ehtimalı bir o qədər yüksəkdir. Nitq bazasına daxil olan material universal deyil, həmişə fərdi olur və konkret natiqlərin tələffüz xüsusiyyətlərini əks etdirir. Əksər dillər üzrə istifadə olunan nitqin mətn üzrə eksperimental sintezatoru konkatenativ sintezatorlar üçün ənənəvi modullardan ibarətdir [2; 3].

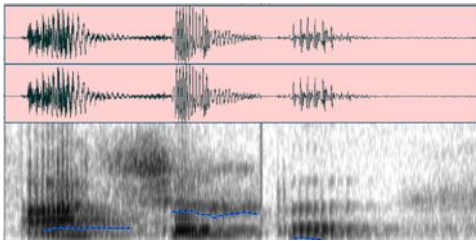
Adətən nitqin tanınması ilə bağlı tədqiqatlarda nitq signalının elementi kimi səs qəbul olunaraq fonemlərin bir-birinin ardınca – ardıcılıqla dərk olunduğu, tanındığı qeyd olunur. Nitqin qavranılması ilə bağlı tədqiqatlar, perseptiv araşdırmalar göstərir ki, fonemlər paralel şəkildə də qavranıla bilər və fonetik element kimi

məhz heca qəbul olunmalıdır. Nitq sintezi zamanı məhz hecanın xüsusi fonetik element olduğunu diqqət mərkəzində saxlamaq mühüm şərtidir. Bu zaman diqqət mərkəzində saxlanılmalı olan məsələlərdən biri də tənəffüs zamanı bir nəfəsvermədə tələffüz olunan hecaların sayı məsələsidir.

Nitq sintezi zamanı ritmik mənərə tonal vasitələrlə ifadə olunmalı, bunun üçün uyğun üsullar ilənib hazırlanmalıdır. Eksperiment zamanı prosodiya ilə bağlı problemləri aradan qaldırmaq məqsədi ilə səsin stasionar hissəsinin “klonlaşdırılması”ndan və pauzadan istifadə etdik. Vurgulu hecada saitin daha “qüvvətli” səslənməsinə nail olmaq üçün “klonlaşdırma”, eləcə də pauzadan istifadə sintezin keyfiyyətinə öz təsirini göstərir (Şəkil 1).



(açıqdır – təbii nitq)



(açıqdır - sintez)

Şəkil 1. Təbii nitqin sintezi

Qeyd edək ki, heca sərhədini müəyyən etmək məqsədilə samitin daxilində “klonlaşdırma” əməliyyatı aparılması da müsbət nəticə verir.

Sintezdə nailiyyət əldə etmək üçün nitqin fizioloji xüsusiyyətləri diqqət mərkəzində saxlanılmalıdır. Nəfəsalma, nəfəsvermə və onlar arasındakı pauzadan ibarət olan tənəffüs sikli 2,5-7 s. təşkil edir. Bir çox insanlarda nəfəsalmanın davamlılığı nəfəsvermədən daha azdır – bu mərhələ daha qısaadır. Pauzanın davamlılığı fərqlidir – hətta olmaya bilər.

Nəzərə almaq lazımdır ki, nitq əksər dünya dillərində nəfəsvermə zamanı baş verir. Nitq signalının seqmentlərə ayrıldığı parçada (bizim təcrübədə hecada seqmentləşmə samit və sait kimi yox, sait və sait olmayan) sait olmayan iki samit, samit və pauza, ya da pauza ola bilər. Prosodiyanın təmin olunması üçün sintezdə bu məqamlarda pauza xüsusi vasitə kimi əhəmiyyətli rola malikdir. Stasionar sahələrin davamlılığı (ölçüsü) keçid hissələrdən böyük olsaydı, həqiqətən də fonemlər ardıcıl dərk olunardı – eksperimental tədqiqat və təcrübə göstərir ki, adətən keçid hissələr daha “uzun”dur. Təbii nitq kəsiyində bəzən müəyyən hissədə müxtəlif fonemlərə məxsus xüsusiyyətlər “qarışıq” şəkildə müşahidə olunur – əvvəlki və sonrakı fonemin sərhədini dəqiq müəyyən etmək çətin olur. Bu hissələrdə samitə aid fiziki parametrlərin mövcudluğu onun samitə aid olunması fikrini doğursa da, bu hissə kəsildikdə saitin müəyyən olunması çətinləşir. Bu tipli eksperimentlərə əsasən, sait və samitin ardıcıl yox, paralel şəkildə dərk olunması ehtimal olunur.

Saitlərin uzunluğu samitin növündən də asılıdır. Novlu samitdən əvvəl sait daha uzundur, nəinki kipləşən-partlayan samitdən. Nəzərə almaq lazımdır ki, sözdə hecalar artdıqca, nitq prosesi ilə bağlı olaraq fizioloji xüsusiyyətlərin təsiri ilə saitin ölçüsü qısalar. Kar və ya cingiltili samitdən əvvəl olmasından asılı olaraq, saitin uzunluğu fərqli olur: kar samitdən əvvəl saitin uzunluğu cingiltilidən əvvəl olanla müqayisədə uzunluğunun 1/2, 1/3-nə bərabərdir. Deməli, sintez zamanı samitin kar və ya cingiltili olduğunu qabartmaq üçün ondan əvvəlki saitin uzunluğu əhəmiyyətli amil kimi istifadə oluna bilər. Sait səslərin uzunluğunun adi halda 150-350 ms. arası olduğunu nəzərə almaqla qeyd edə bilərik ki, sintez zamanı uzunluğu 50 ms. olan saitin 100 ms.yə çatdırılması onun daha yüksək səs kimi (təxminən 2-3 dB.) dərk olunmasına şərait yaradır. Maraqlıdır ki, 100-200 ms. arası dəyişmə bu effekti vermir.

Aparılan eksperimentlərin nəticəsinə görə, hecanın “davamlılığı” (uzunluğu - ölçüsü) saitin uzun tələffüzü hesabına təmin olunur – balans saithlərin ölçüsü vasitəsilə yaradılır. Bu eksperimentlərin nəticəsinə görə, vurğu da məhz sait səsin uzunluğu ilə təmin olunur. Nəzərə almaq lazımdır ki, hər dil üçün vurğunu müəyyən edən fiziki parametrlər eyni deyil. Bu kəmiyyət vurğusuna malik dillər üçün səciyyəvidir, yeri gəlmişkən, bu məsələ Azərbaycan dili üçün özünü doğrultmur. Belə ki, bir sıra dillərdə başqalarından uzun heca vurğulu kimi dərk olunur. Bu vasitə ilə sintez zamanı cümlədə məntiqi vurğunun təmin

olunmasına da nail olmaq olar. Qeyd edək ki, samitin uzunluğu sözün sərhədini bildirir və bu amil sintez zamanı təbii nitqə yaxın nəticə əldə etmək üçün xüsusi əhəmiyyət daşıyan amillərdən biri kimi istifadə oluna bilər.

Nitq mürəkkəb və çəxtərəfli prosesin nəticəsi olduğundan onun təhlili zamanı fizioloji, psixoloji, koqnitiv, akustik xüsusiyyətlərin nəzərə alınması və kompleks şəkildə təhlili vacibdir. Əks halda nitq aktının mexanizmi və nitq signalının parametrləri ilə bağlı dolğun təsəvvür əldə etmək mümkün deyil. İ.M. Seçenovun da qeyd etdiyi kimi, biz səssiz düşünəndə belə, danışanda olduğu kimi, dilimizdə biotoklar güclənir. Bunu xüsusi qurğular vasitəsilə qeydə almaq mümkündür. Lakin bəzən müəyyən yazını xüsusi diqqətlə “ürəyimizdə” oxuduqda da bizdən asılı olmayaraq, nəinki dil, hətta dodaqlar belə tərpənir... “Daxili nitqin refleksləri funksional baxımdan dəyişilmiş adi nitq refleksləridir” [4]. Daxili nitqin mexanizmi sona qədər tədqiq olunmasa da, aparılan çoxsaylı eksperimentlərin nəticələri daxili nitqdə nitq analizatorlarının iştirakını sübut edir [5]. Alınan nəticələrə görə, çətin məsələləri düşünən zaman artikulyasiya hərəkətləri daha qabarıq şəkildə ifadə olunur. Daktil nitqi (lalların barmaq işarələri vasitəsilə həyata keçirdiyi nitq) mənimsəmiş sağlam insanlarda daxili nitq zamanı əvvəlcə dil, daha sonra barmaq əzələlərində yüksək elektroaktivlik müşahidə olunur.

### III. NİTQ SİNTEZİNDƏ NİTQİN PROSODİK VASİTƏLƏRİNDƏN İSTİFADƏNİN ÜSTÜNLÜYÜ

Nitq axınının fonetik (seqmental və suprasedimental) təşkili haqqında məlumatlara kifayət qədər diqqət yetirilməməsi bəzən signalın təbiiliyinin pozulmasına səbəb olur. Bunu makrosintez nümunələrində təbii nitq signalınının kifayət qədər uzadılmış fraqmentlərində asanlıqla görmək olar. Makrosintez dəmir yolu vağzallarında elanlarda, danışan saatlarda və oyun proqramlarında istifadə olunur. Bu pozuntular sözlərin qovşağındakı samit birləşmələri, nitqin ritmik təşkilinin pozulması və ya qeyri-təbii intonasiyada ifadə olunur ki, bu da dil norması nöqtəyindən qəbul edilməz hesab olunur və sintezin keyfiyyətini müəyyən edən amillərdəndir.

Nitq signalınının rəqəmsal şəkildə emalı əsas tonun tezliyinin və signal fraqmentlərinin müddətinin dəyişdirilməsi ilə sintezdə qarşıya çıxan müəyyən problemləri həll etməyə imkan verir. Səs mənbəyinin iştirakı olmadan həyata keçirilən nitq signalınının fraqmentinin müddətini dəyişdirmək (məsələn, samitlərin parametrlərində dəyişiklik, fiziki fasilələr) kifayət qədər sadə üsullarla həll olunur. Kipləşən-partlayan samitlərdə pauzanın və ya küy hissəsinin müddətini dəyişdirmək – müddəti artırmaq üçün sıfır amplitudalı bir hissə əlavə etməklə və ya onu azaltmaq üçün küy və ya fasilənin bir hissəsini çıxarmaqla nəticə əldə etmək mümkündür. Afrikatların müddətini dəyişdirmək bir qədər daha çətindir. Təcrübələr göstərir ki, yalnız pauza və ya yalnız səs-küyün müddətini dəyişdirmək orijinal fonetik

kontekstdə səs qavrayış keyfiyyətinin pozulmasına səbəb olur. Buna görə də, affrikatın müddətində dəyişiklik pauza və səs-küyün müddətinin mütənasib dəyişməsi ilə əldə edilə bilər. Təbii ki, kar samitlərin müddətini manipulyasiya edərkən, bu samitlərin partlayış sahəsinə və ətrafdakı səslərə bitişik ərazilərə müdaxilə etmək yol verilməzdir.

Kar samitlərə nisbətən sonorların və saitlərin müddətini dəyişdirmək daha çətindir. Belə ki, səs mənbəyinin iştirakı ilə yaranan səslərin müddəti üçün ölçü vahidi yalnız saniyələrlə (yaxud daha əlverişlisi millisaniyələrlə) ölçülən müddətlə deyil, həm də səs mənbəyinin dövrlərindəki müddəti ilə (bu cür səslərin daxili təşkili ilə) müəyyən edilən əsas tezlik hesablanır, dəyişə bilər. Bu baxımdan, belə səslərin müddətini dəyişdirmək yalnız əsas tezliyin dövrlərini manipulyasiya etməklə (lazımı sayda dövrləri silmək və ya köçürməklə) əldə edilə bilər.

Saiti əmələ gətirən dövrlərin seçilməsinə obyektiv olaraq əlavə məhdudiyətlər qoyulmaqla, onun müddətinin dəyişməsinə nail olmaq olar, lakin nəzərə almaq lazımdır ki, bu cür manipulyasiyalara ən dözümlü hissə saitın stasionar bölməsidir. Odur ki, manipulyasiya zamanı bitişik dövrlərdən – koartikulyasiya hissələrindən istifadə arzuolunmazdır, əməliyyata məruz qalan sahələr arasındakı məsafə ən azı bir dövr olmalıdır.

Saitlərdə və sonor səslərdə əsas tonun tezliyində xüsusi dəyişiklik baş verir ki, bu da nitqin melodik tərtibatını müəyyənləşdirir. Bu baxımdan, konkret səs müddətinin dəyişməsi nitqin melodik tərtibatına təsir etməməli və ya melodik

tərtibatın korreksiyası (əsas ton tezliyinin modifikasiyası) ilə müşayiət olunmalıdır. Əks halda bu ona gətirib çıxarır ki, səslərin müddətinin dəyişməsi, bir qayda olaraq, nitqin və ya onun hissəsinin melodik tərtibatının dəyişdirilməsi ilə nəticələnir. Bununla belə, əsas tonun tezliyinin dəyişməsi həm də əsas tonun dövrlərinin müddətinin dəyişməsi deməkdir ki, bu da təbii olaraq siqnalın dəyişən fraqmentinin ümumi müddətinin dəyişməsinə gətirib çıxarır. Fonetik təcrübələrdə və yüksək keyfiyyətli avtomatik nitq sintezi sistemlərinin işlənilib hazırlanmasında əsas ton tezliyinin modifikasiyasından istifadənin əvəzsiz şərti onun qavrayış keyfiyyətinin və təbiiyyətinin mütləq şəkildə qorunub saxlanmasını təmin etməkdir. Problem ondadır ki, əsas tonun tezliyindəki təbii dəyişiklik, bir qayda olaraq, saitlin linqvistik baxımdan tanınmasına cavabdeh olan formantların mövqeyinə təsir etmir, əksinə, bəzi harmonikalarda səsin fərdi tembrini ilə bağlıdır və qalan harmonikaların yerini dəyişir. Odur ki, spektrin bütün digər komponentlərini eyni yerdə saxlamaqla yalnız əsas tezliyin dəyişdirilməsi qeyri-təbii siqnal təəssüratı yaradır.

Sintezin təbii nitqə daha yaxın olması üçün təbii nitqin komponentləri maksimal şəkildə sintezdə öz əksini tapmalı, həm seqment, həm də superseqment səviyyəyə aid elementlər korpusun təşkilində olduğu kimi, sintez prosesində də nəzərə alınmalıdır. Adətən sintezin “süni” səslənməsi superseqment vasitələrin – prosodik xüsusiyyətlərin dəqiqliklə ifadə oluna bilməməsindən qaynaqlanır. Bir çox hallarda, hətta əhatəli korpusun mövcudluğu

belə, sintezdə prosodiya məsələsini həll etmək üçün yetərli olmur.

İlk dəfə İ.Mattinqli qaydalar üzrə prosodik sintezatorun layihəsini təqdim etmiş [6], 2000-ci ildə isə Burkhardt [7] formant sintezindən istifadə edərək müxtəlif emosional kateqoriyalar üçün nitqin əsas akustik parametrlərini müəyyən etmişdir. Burkhardtın tədqiqatının məqsədi neytral şəkildə ifadə edilən nitqin akustik parametrlərini dəyişdirərək emosional qavrayışa təsir edən akustik xüsusiyyətləri müəyyən etmək idi. Onun təcrübələrinə görə, emosional nitqin formant sintezində əsas ton, onun orta qiyməti və dəyişmə diapazonu, nitqin sürəti, fonasiya, sait tərifinin dəqiqliyi kimi parametrlər həlledici əhəmiyyət kəsb edir.

Keyfiyyətli sintezin ilkin şərtlərindən biri bunun üçün qurulmuş korpusun mövcudluğudur. Müasir dövrdə geniş şəkildə istifadə olunan difonik sintez modeli səs bazasının formalaşması üçün müəyyən prinsipləri nəzərdə tutur [8]. Birinci mərhələdə, bütün mümkün birləşmələri sintez üçün istifadə olunan difonların sayını məhdudlaşdıran vahidlərin siyahısını müəyyən etmək, həm kəmiyyət, həm də keyfiyyət baxımından əsas allofonların siyahısının tərtib olunması, sait və samitlərin bəzi kombinator variantları, eləcə də saitlərin mövqe variantları nəzərə alınaraq korpus yaradılmalıdır. Bundan sonra həm sözün daxilində, həm də sözlərin qovşağında yaranan və nəzəri cəhətdən mümkün C + U, C + C, U + C, U + U birləşmələr nəzərə alınmaqla xüsusi difonik baza tərtib edilməlidir.

## NƏTİCƏ

Nitq signalının işlənməsi zamanı fonemlərin koartikulyasiya, birləşmə və ritmik mövqə nəticəsində yaranan məcburi və potensial mümkün modifikasiyaları nəzərə alınmalıdır. Bu cür sistemləri öyrətmək üçün müəyyən bir dilin fonemlərinin bütün mümkün mövqə və kombinator variantlarını özündə birləşdirən ifadələr və ya ifadələr korpusunun olması əsas şərtidir. Bu prinsipdən istifadə edən sistemlər müəyyən bir lüğətlə məhdudlaşmır, buna görə də əlavə təlim olmadan onlara istənilən sayda yeni söz əlavə edilə bilər. Konkatenativ üsulla sintezin uğurlu alınmasının ilkin şərti verilənlər bazasının böyüklüyü və əhatəliliyi olsa da, təbii nitqə yaxın sintezə nail olmaq üçün müasir eksperimental fonetik araşdırmalardan əldə olunan nəticələrin tətbiqi ilə “klonlaşdırma” üsulundan istifadə daha yaxşı nəticəyə nail olmaq üçün faydalı ola bilər.

## İSTİNADLAR

- [1] Зиндер Л.Р. Общая фонетика. – м.,1979.
- [2] Krivnova O.F. Automatic synthesis of Russian speech // Proceedings of the XIV International Congress of Phonetic Sciences, Vol.1, pp. 507–510, San Francisco, 1999.
- [3] Лобанов Б.М., Карневская Е.Б., Левковская Т.В. Синтезатор речи по тексту как компьютерное средство «клонирования» персонального голоса // Тр. Международной конференции Диалог-2001 / М., 2001. С. 265-272.
- [4] Сеченов И. М. Рефлексы головного мозга. 1942.
- [5] Соколов А. Н. Внутренняя речь и мышление. М., Искусство, 1968.
- [6] Klatt D. H. Review of text-to-speech conversion for English // The Journal of the Acoustical Society of America. 1987. Vol. 82. No 3. P. 737–793.
- [7] Burkhardt F., Sendlmeier W. F. Verification of acoustical correlates of emotional speech using formant-synthesis // ISCA Tutorial and Research Workshop (ITRW) on speech and emotion. 2000.
- [8] Бондарко Л.В., Вербицкая Л.А., Гордина М.В. Основы общей фонетики. Л., 1983.