

АНАЛИЗ НА ВЛИЯНИЕТО НА ВРЕМЕТРАЕНЕТО НА ФАЗАТА НА ОТВОРЕН ГЛОТИС ВЪРХУ СПЕКТРАЛНИТЕ ХАРАКТЕРИСТИКИ НА ПРОЦЕСА НА ФОНАЦИЯ

Дамян Дамянов, Васил Гълъбов

Резюме: Процесът на фонация, като част от речеобразуването, се описва от множество модели на генериране на реч. Повечето методи за определяне на параметрите на филтъра изискват приемането на хипотезата за линеен чисто-тоположен филтър. Един от разпространените в теорията и практиката модели е този на Фант, който приема, че процесът е от типа "импулсен източник - филтър". Приемането на чисто-тоположен модел на филтъра е исторически обосновано от желанието да се постигне високо бързодействие в някогашните системи за обработка на речеви сигнали в реално време. В статията се демонстрира, че моделът на Фант обяснява някои феномени на речеобразуването, наблюдавани при промяна на моментното психофизиологическо състояние на човека. Такива са изменението на отношенията на амплитудите на формантните честоти (по-дрезгав или по-звънтящ глас) и др.

Ключови думи: системи с речева комуникация, модел на Фант, фаза на отворен глотис, човеко-машинни системи

ON THE IMPACT OF DURATION OF THE PHASE OF OPEN GLOTTIS ON THE SPECTRAL CHARACTERISTICS OF THE PHONATION PROCESS

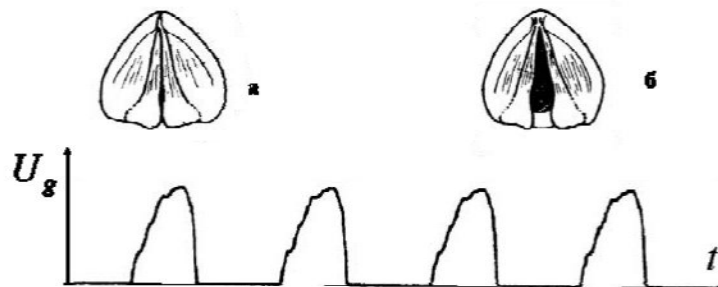
Damyam Damyanov, Vassil Galabov

Abstract: The phonation process, as a part of speech production, is described by many models for speech generation. Most of the methods for estimation of the coefficients of the filter, assume an all-pole filter. One of the common models in theory and praxis is the model of Fant, which utilizes the "pulse driver - filter" model. The assumption for all-pole filter model has historical background, dating back from the time when speech processing systems had to work in real-time and the processing power was very limited. As it is stated in this paper, the model of Fant explains some of the phenomena of speech production, which can be experienced during change of momentary psychophysical condition of the human being. Such are the ratio of the amplitudes of the formant frequencies (hoarser or more clinking voice).

Keywords: speech communication systems, model of Fant, phase of open glottis, man-machine systems

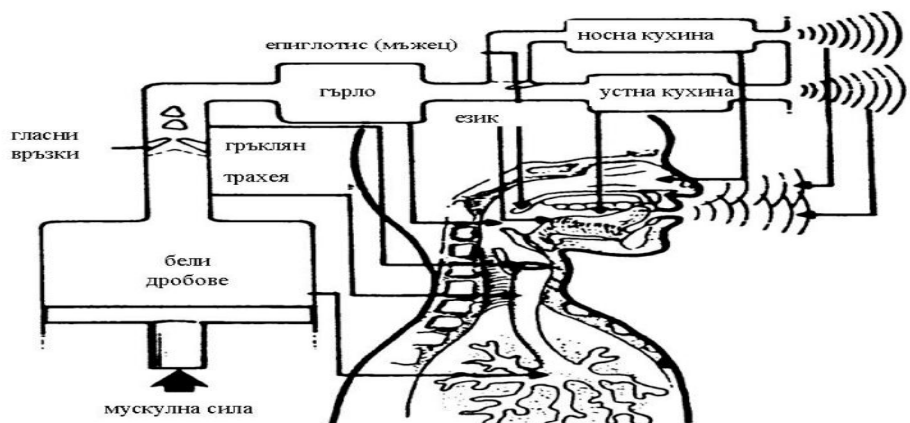
1. ВЪВЕДЕНИЕ

Фонацията е процесът, който се извършва в средната част на ларинкса и озвучава издишваната струя въздух. Като последица от този процес потокът на звукова скорост непосредствено след гласната цепка проявява периодичен характер (фиг.1), който определя основният тон на речевия сигнал [10]. Промеждият напрежението на гласните гънки и налягането в процеса на речта, човек получава необходимата честота на основния тон осъществяване на фонацията на гласните и на звучните съгласни. При образуването на беззвучните съгласни гласните гънки не се приближават една до друга и положението им е като при физиологично дишане [1,8]. В този случай постъпващият въздушен поток от белите дробове се модулира от образуването на турбуленции и от затварянето на вокалния тракт. Спектрите на резултиращите звукови източници се модифицират от резонансни явления, честотата на които зависи от променящата се форма на гърлото и устната кухина, от разположението на езика и др.



Фиг.1 Поток на звукова скорост непосредствено след гласната цепка при фонацията на гласни и звучни съгласни. Положение на гласните гънки: а - фонацията на гласни и звучни съгласни, б - физиологично дишане

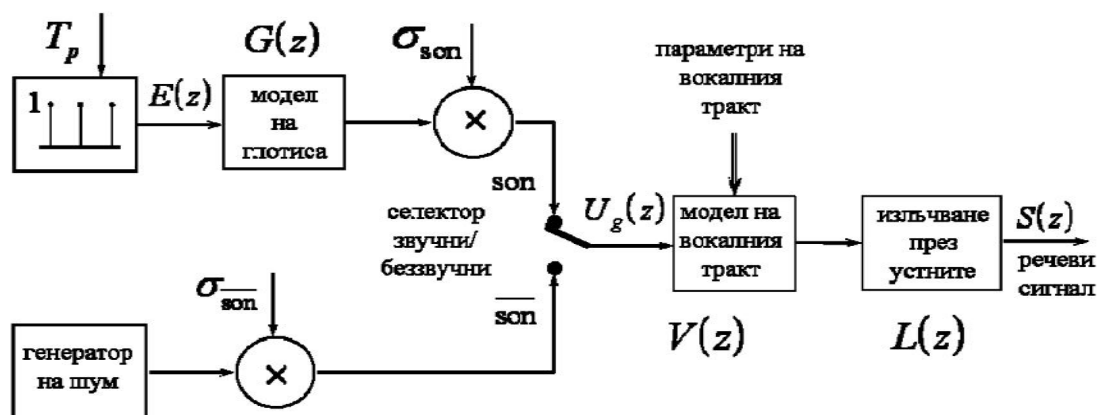
Целта на моделирането на процесите на генериране на речевия сигнал е не само да бъдат изследвани неговите свойства и особености, а и да се решат задачите за неговото ефективно кодиране и пренос, за автоматичното синтезиране и разпознаване на реч, за психофизиологични и биометрични приложения и за много други. Във всеки такъв модел [5], в една или друга степен, са отразени главните компоненти на артикулационния тракт от гледна точка на законите на акустиката (фиг.2)



Фиг.2 Основни акустични компоненти на артикулационния тракт

Към тези компоненти спадат възбуждането на гласните връзки, променящата се във времето форма на вокалния тракт, излъчването от устните и т.н. Резонансните явления при звукооформянето на практика са повлияни от загуби в стените на тракта в зависимост от топлопроводимостта, еластичността и триенето. Значително влияние оказва и носната кухина с поглъщането на определени честотни компоненти от спектъра (антирезонансни явления). Този сложен характер на процеса на речеобразуване позволява синтезиране на различни по структура и по степен на адекватност негови модели. Интересно е, че въпреки очевидната сложност на проблема, той е намерил едно сравнително тривиално решение, предложено от *Fant* [4], което се е утвърдило в практиката независимо от конкретното предназначение на съответната система.

Класическият модел на *Fant* отделя възбуждането от оформянето на отделните звукови компоненти на речта, което позволява приблизително третиране на речевия апарат като линейна система. В съвременните системи този модел е намерил своята дискретна реализация (фиг.3). За повечето практически приложения може да се приеме, че речевият сигнал се състои от чисто звучни или беззвучни компоненти със съответна амплитуда. Поради постепенната промяна на възбуждането и звукооформянето параметричният анализ може да се извършва поинтервално, като в рамките на 10-40 ms моделът се приема за квазиинвариантна във времето линейна система [4,7,9].



Фиг.3 Дискретизиран модел на *Fant*.

2. ПРЕДПОСТАВКА ЗА АНАЛИТИЧНО ИЗСЛЕДВАНЕ

Голяма част от методите за изследване са заимствани от медицината, както обикновено, за да се постигне по-голяма точност или да се осъществи по-лесно измерването, не се работи директно с търсената величина, а с друга, зависима от нея. Поради сложността на провеждане на ларингоскопски и ларингоендоскопски изследвания и най-вече поради необходимостта от медицинска интервенция и поради значителното въздействие на тези методи върху цялостния процес на речеобразуване са разработени и редица индиректни методи, на основата на електромиографията - регистрират се сигналите на мускулната активност в гръкляна по време на говорене, глотографията - регистрира се импеданса на

шията в ранината на глотиса, който в голяма степен зависи от сечението на въздушния поток, преминаващ през глотиса. Въпреки многобройните изследвания, към момента липсва единна теория, която да обяснява процеса на тези изменения в спектралните характеристики на речевия сигнал. Повечето автори търсят причината за тези феномени в психо-физиологично обусловена промяна на моментната форма артикулационния тракт. Моделът на *Fant* позволява по метода на линейното предсказване да се извърши реконструкция (с някаква степен на приближение) както на моментната форма на тракта, така и на възбуждащия поток на звукова скорост след глотиса. Очаква се, че моделирането на речевия тракт осигурява условия за получаване на релевантна информация за функционирането на facialната мускулатура и съответните участъци на нервната система, които са крайно чувствителни към моментното психо-физиологично състояние на индивида [2,3]. Но практиката показва, че флукуациите на речевия тракт вследствие психо-физиологично повлияване на функционирането на facialната мускулатура в повечето случаи са твърде незначителни и са под точността, с която линейният модел позволява реконструкция на тракта, като го апроксимира с каскада от съосни цилиндрични участъци с еднаква дължина и константно сечение. Това налага усложняване на алгоритмите, като допълнителна информация се извлича от модела на потока на звукова скорост след глотиса. Този поток носи информация за моментното психо-физиологично състояние по два канала - първо чрез цялостния процес на речеобразуване, ангажиращ висшата нервна система, и второ чрез функционалното състояние на вегетативната нервна система, включително и общия мускулен тонус [2,6]. За извличане на тази информация се налага използване на сравнително сложни модели, описващи напрегнатостта на гласните връзки, механизма на тяхното опъване и движение, изменение на налягането на постъпващия от белите дробове въздушен поток и др.

В настоящото изследване ще покажем, че моделът на *Fant* всъщност позволява описването на психо-физиологично обосноващите промени в спектралните характеристики на речевия сигнал без да се налага използването на допълнителни модели. За целта е достатъчно да се анализират взаимовръзките на основни параметри на възбуждащия импулс на източника с честотните характеристики на филтъра. В съществуващата практика тези взаимовръзки не се имат предвид и източникът и филтърът се разглеждат поотделно. За по-голяма разбираемост в по-нататъчните разглеждания ще бъдат направени следните опростявания, които не променят същността на изводите:

- влиянието на предавателната функция на глотиса и на излъчването през устните ще бъдат пренебрегнати, тъй като не влияят на формантните честоти, а само на тяхното затихване;
- възбуждането от източника ще се разглежда като последователност от правоъгълни импулси;
- затихването на формантните честоти ще се пренебрегне, т.е. полюсите на филтъра на вокалния тракт ще лежат на единичната окръжност;

3. ВЛИЯНИЕ НА ВРЕМЕТРАЕНЕТО НА ФАЗАТА НА ОТВОРЕН ГЛОТИС ПРИ ПОВЕЧЕ ОТ ЕДНА ФОРМАНТНА ЧЕСТОТА

Приемаме, че моделът на вокалния тракт е чисто полюсен от четвърти ред и полюсите лежат върху единичната окръжност:

$$H(s) = \frac{1}{(s^2 + \omega_1^2)(s^2 + \omega_2^2)} \quad (1)$$

При възбуждане на филтъра с последователност от правоъгълни импулси от типа:

$$e(t) = \begin{cases} 1, & (m-1)T_g \leq t < t_{open_glottis} + (m-1)T_g \\ 0, & t_{open_glottis} + (m-1)T_g \leq t < mT_g \end{cases} \quad (2)$$

$$m = 1, N_{imp} - 1$$

където T_g е периодът на основния тон, а $t_{open_glottis}$ е продължителността на фазата на отворен глотис, изходният сигнал на филтъра за първия период на основния тон ($m=1$) ще съдържа два форманта с кръгови честоти ω_1 и ω_2 . Ако приемем, че $\omega_2 = k_2\omega_1$ и $k_2 > 1$, то компонентите на сигнала, съответстващи на отделните форманти, във фазата на отворен глотис ще бъдат:

$$s_{open_glottis}^{1F}(t) = AO_{open_glottis}^{1F} - A_{open_glottis}^{1F} \sin(\omega_1 t + \varphi_{open_glottis}^{1F}) \quad (3)$$

$$s_{open_glottis}^{2F}(t) = AO_{open_glottis}^{2F} - A_{open_glottis}^{2F} \sin(\omega_2 t + \varphi_{open_glottis}^{2F}) \quad (4)$$

където:

$AO_{open_glottis}^{1F} = \frac{1}{(k_2^2 - 1)\omega_1^4}$ е постояннотоковите съставки на сигнала, описващ формата на първата форманта при отворен глотис,

$AO_{open_glottis}^{2F} = \frac{1}{-k_2^2(k_2^2 - 1)\omega_1^4}$ е постояннотоковите съставки на сигнала, описващ формата на втората форманта при отворен глотис,

$A_{open_glottis}^{1F} = \frac{1}{(k_2^2 - 1)\omega_1^4}$ е амплитудата на сигнала, описващ формата на първата форманта при отворен глотис,

$A_{open_glottis}^{2F} = \frac{1}{k_2^2(k_2^2 - 1)\omega_1^4}$, е амплитудата на сигнала, описващ формата на втората форманта при отворен глотис,

$\varphi_{open_glottis}^{1F} = \frac{\pi}{2}$ е фазовото отместване на сигнала, описващ формата на първата форманта при отворен глотис,

$\varphi_{open_glottis}^{2F} = \frac{3\pi}{2}$ е фазовото отместване на сигнала, описващ формата на втората форманта при отворен глотис,
 $\omega_1 = 2\pi f_1$ и $\omega_2 = 2\pi f_2$ са кръговите честоти, съответстваща на формантите честоти f_1 и f_2 .

Във фазата на затворен глотис върху параметрите на сигналните компоненти влияние ще оказват безразмерните коефициенти, пропорционални на времетраенето на фазата на отворен глотис към периода на съответната формантна честота:

$$r_{\omega_1 t_{open_glottis}} = \omega_1 t_{open_glottis} \text{ и } r_{\omega_2 t_{open_glottis}} = \omega_2 t_{open_glottis} = k_2 r_{\omega_1 t_{open_glottis}} \quad (5)$$

Това може да бъде представено чрез безразмерния коефициент на отношението на амплитудите на двете формантни честоти в сигнал:

$$r_{A^{2F} A^{1F}} = \frac{A^{2F}}{A^{1F}} \quad (6)$$

Във фазата на отворен глотис този коефициент се определя само от параметрите на филтъра:

$$r_{A^{2F}_{open_glottis} A^{1F}_{open_glottis}} = \frac{1}{k_2} \quad (7)$$

Във фаза на затворен глотис обаче, това отношение, освен от параметрите на филтъра, зависи по сложен начин и от продължителността на възбуждащата фаза, т.е. на отворен глотис:

$$r_{A^{2F}_{open_glottis} A^{1F}_{open_glottis}} = \frac{1}{k_2} \frac{\left| \sin\left(\frac{k_2 r_{\omega_1 t_{open_glottis}}}{2}\right) \right|}{\left| \sin\left(\frac{r_{\omega_1 t_{open_glottis}}}{2}\right) \right|} \quad (8)$$

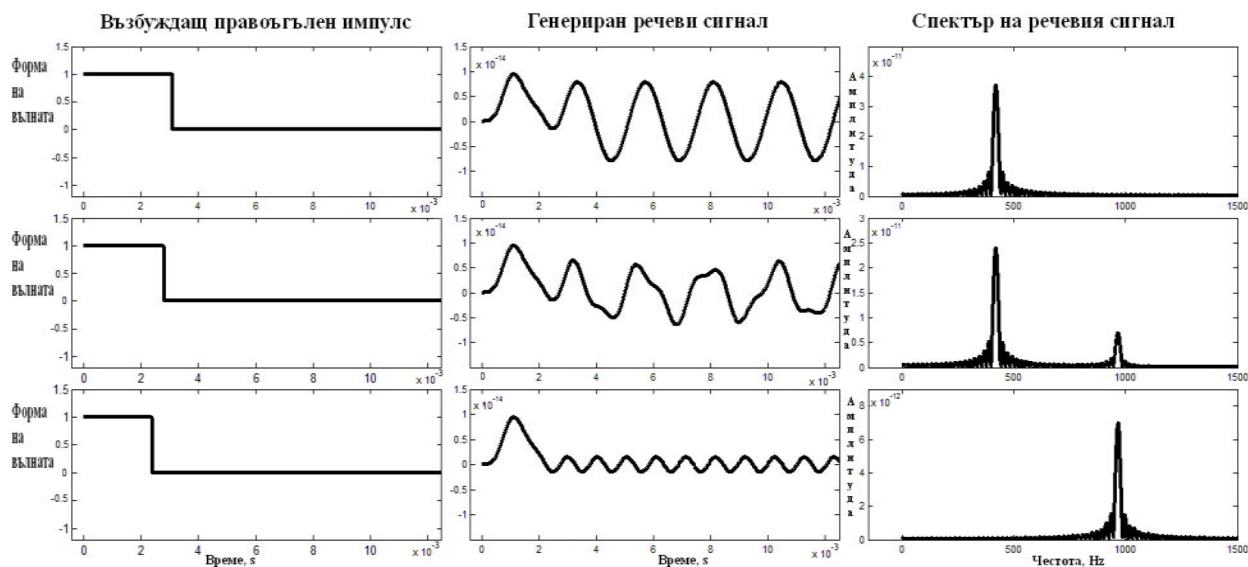
На практика това означава, че промяната на продължителността на фазата на отворен глотис може съществено да изменя предопределената от геометрията на вокалния тракт констелация на формантните честоти в сигнала. Това влияние ясно личи от следния числов пример, с типични за реален речеви сигнал стойности на параметрите на модела:

- първа формантна честота $f_1 = 420 \text{ Hz}$;
- втора формантна честота $f_2 = 966 \text{ Hz}$, която съответства на $k_2 = \frac{f_2}{f_1} = 2.3$;
- номинална продължителност на фазата на отворен глотис $t_{open_glottis} = 2.8 \text{ ms}$;
- флукутация на номиналната продължителност на фазата на отворен глотис $\Delta t_{open_glottis} = \pm 0.4 \text{ ms}$;

Във фазата на отворен глотис отношението на амплитудите зависи само от параметрите на филтъра и ще бъде:

$$r_{A_{open_glottis}^{2F} A_{open_glottis}^{1F}} = \frac{A_{open_glottis}^{2F}}{A_{open_glottis}^{1F}} = k_2^{-2} \approx 0.189 \quad (9)$$

Във фазата на затворен глотис на номиналната продължителност на $t_{open_glottis} = 2.8 \text{ ms}$ този коефициент ще бъде $r_{A_{open_glottis}^{2F} A_{open_glottis}^{1F}} \approx 0.288$. При намаляване на продължителността на отворената фаза на глотиса с 4 ms коефициентът ще нарасне повече от 20 пъти до $r_{A_{open_glottis}^{2F} A_{open_glottis}^{1F}} \approx 6.33$, и съответно при увеличаване на продължителността с 4 ms - ще намалее повече от 20 пъти до $r_{A_{open_glottis}^{2F} A_{open_glottis}^{1F}} \approx 0.061$. На фиг.4 за трите случая са показани възбуждащия правоъгълен импулс с продължителност, съответстваща на продължителността на отворен глотис, генерирания от филтъра сигнал и неговия честотен спектър.



Фиг.4 Влияние на флукуацията на продължителността на продължителността на фазата на отворен глотис върху констелацията на формантните спектри на сигнала

4. ИЗВОДИ

Моделът на *Fant* дава отлични резултати в повечето случаи на използване на модела - основно при системите за анализ, синтез, кодиране и пренасяне на речеви сигнали. При използването обаче на речеви сигнали за медицински, психофизиологични, биометрични и др. цели е прието да се смята, че моделът на *Fant* не е достатъчно релевантен и се налага използване на значително по-усложнени модели и допълнителни информационни източници. Напоследък проблемът става още по-актуален с постоянно увеличаващите се изисквания към качеството на системите за обработка и пренос на речеви сигнали и използването им в мобилни устройства.

Направеното изследване показва, че моделът може да бъде направен значително по-ефективен без да се усложнява, а като се използва кумулативния ефект на едновременно разглеждане на някои процеси, протичащи в източника и филтъра.

5. БЛАГОДАРНОСТИ

Авторите изказват своята благодарност към НИС при ТУ-София за финансовата подкрепа по проект № 122ПД0014-08, в рамките на който е проведено настоящето изследване.

ЛИТЕРАТУРА

- [1] Тилков Д., Бояджиев Т. (1990), *Българска фонетика*, Наука и изкуство. София, 1990
- [2] Birbaumer N., Schmitt R.F. (1996), *Biologosche Psychologie*, Springer, Berlin, 1996
- [3] Ekman, P., Friesen W. (1978), *The Facial Action Coding System*, Consulting Psychologists Press, San Francisco, CA, 1978
- [4] Fant G. (1990), *Acoustic Theory of Speech Production*, Mouton & Co., Hauge, 1990
- [5] Flanagan J. (1992), *Speech Analysis, Synthesis and Perception*, Springer, Berlin, 1992
- [6] Gazzaniga M. (1995), *The Cognitive Neuroscience*, MIT Press, New York, 1995
- [7] Hayes M. (1999), *Schaums's Outline of Theory and Problems of Digital Signal Processing*, Singapore, McGraw-Hill, 1999
- [8] Pickett J.M.(1989), *The Sounds of speech communication*, University Park Press, Baltimore, 1989
- [9] Rabiner L., Schafer R. (1992), *Digital Processing of Speech signals*, Prentice-Hall Inc., Engelwood Cliffs, New Jersey, 1992
- [10] Titze Ingo R. (1984), *Parameterization of the glottal area, glottal flow, and the vocal fold contact area*, JASA, 75(2), February, 1984, pp.570-580

Автори: Дамян Дамянов, асистент, катедра „Автоматизация на Непрекъснатите Производства”, Факултет Автоматика, Технически Университет - София, E-mail address: [damyan.damyamov@fdiba.tu-sofia.bg](mailto:damyam.damyamov@fdiba.tu-sofia.bg), Васил Гълъбов, доц. д-р , катедра „Автоматизация на Непрекъснатите Производства”, Факултет Автоматика, Технически Университет - София, E-mail address: vtg@tu-sofia.bg

Постъпила на 28.04.2012

Рецензент проф. д-р С. Йорданова