

## ОСОБЕНОСТИ НА МОДЕЛА НА ФАНТ ОТ ВТОРИ РЕД ПРИ РЕЧЕОБРАЗУВАНЕТО

Дамян Дамянов, Васил Гълъбов

**Резюме:** Процесът на речеобразуване се описва с модела на Фант, който приема, че процесът е от типа "импулсен източник - филтър". За определяне на параметрите на филтъра съществуват множество методи, повечето от които изискват приемането на хипотезата за линеен чистополюсен филтър. В статията се демонстрира, че моделът на Фант обяснява някои феномени на речеобразуването, като влошена разбираемост например.

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

## CHARACTERISTICS THE MODEL OF FANT OF SECOND ORDER ON SPEECH PRODUCTION

Damyam Damyanov, Vassil Galabov

**Abstract:** The speech production process is described by the model of Fant, which utilizes the "pulse driver - filter" model. When estimating the coefficients of the filter, one can use many methods, most of which assume an all-pole filter. As it is stated in this paper, the model of Fant explains some of the phenomena of speech production as poor intelligibility for example.

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

### 1. ВЪВЕДЕНИЕ

Класическият модел на *Fant* отделя възбуждането от оформянето на отделните звукови компоненти на речта, което позволява приблизително третиране на речевия апарат като линейна система. Белите дробове, създаващи въздушния поток, се представят като постояннотоков източник, изходния сигнал на който се разделя на две части, определящи съответно звучната и беззвучната компонента на сигнала. Импулсният генератор моделира модулирането на въздушния поток от периодичното възбуждане на гласните връзки, а генераторът на шум - от образуването на турбуленции в местата на промяна на сечението на вокалния тракт. Окончателният речеви сигнал се получава, след като сумираните компоненти преминават през филтър с линейна предавателна функция, моделиращ артикулационния тракт.

Съвременното изследване на акустичните явления във фарингеално-оралния тракт започва през 1941 г. с работата на *Chiba u Kajizama* [1]. Фундамента на настоящата теория и практика се изглажда в периода около 1950 г. от *Dunn, Fant, Stevens, House* и др. Системите, обработващи, пренасящи и съхраняващи речеви сигнали, използват разнообразни методи и средства, но всички те са на основата на някакъв модел на функциониране на артикулационния тракт и на потока на звукова скорост след глотиса. 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.

В съвременните системи моделът на *Fant* е намерил своята дискретна реализация (фиг.1). 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 [2,3,4].

## 2. ПРЕДПОСТАВКА ЗА АНАЛИТИЧНО ИЗСЛЕДВАНЕ

Генерирането на звучен речеви сегмент след z-преобразуването на дискретизирания речеви сигнал се описва с уравнението:

$$S(z) = U_g(z)V(z)L(z) \quad (1)$$

$$U_g(z) = \sum_{n=0}^{\infty} (z^{-K})^n G(z) \sigma_{son} = \frac{\sigma_{son}}{1 - z^{-K}} G(z) \quad (2)$$

Приема се, че периодът на дискретизация  $T=1$ , а периодът на основния тон е кратен на него  $T_p = KT$ .

Предавателните функции на модела на глотиса, на модела на вокалния тракт и на излъчването през устните могат да се обединят в обща предавателна функция:

$$H(z) = G(z)V(z)L(z) \quad (3)$$

с което процесът на речеобразуване може да се запише като:

$$S(z) = E(z)H(z) \quad (4)$$

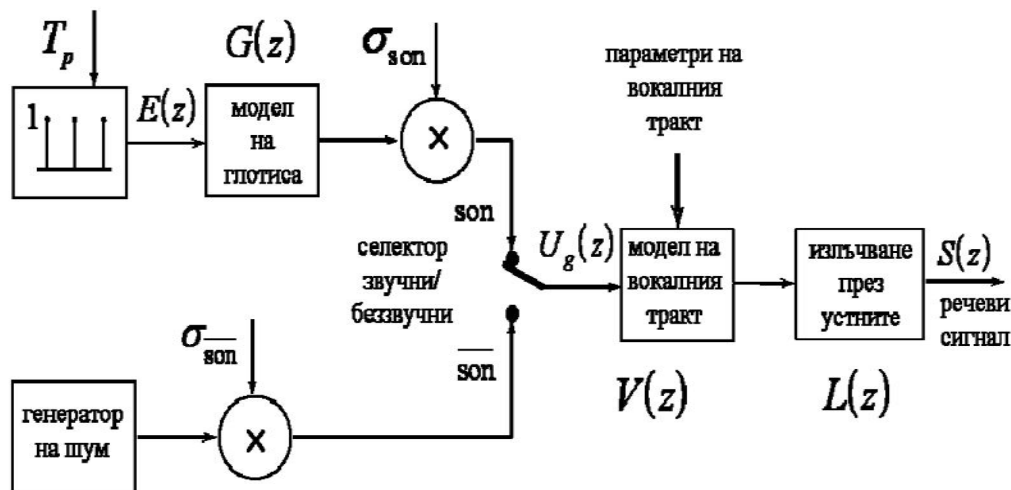
$$S(z) = Z\{s(nT)\}, \quad s(nT) = s(t)|_{t=nT} \quad (5)$$

Където  $E(z)$  е z-преобразуваното импулсно възбуждане на глотиса.

Приемането на определени хипотези за спектралните свойства на глотиса, на вокалния тракт и на излъчването през устните позволява [4,5] позволява в практиката да бъде използван чисто полюсен филтър:

$$H(z) = \frac{1}{1 + \sum_{i=1}^M a_i z^{-i}} \quad (6)$$

за намиране на коефициентите, за което съществуват достатъчно ефективни методи.



Фиг.1 Дискретизиран модел на Fant.

С помощта на процеса на речеобразуване от типа източник-филтър, на основата на модела на Fant, успешно се решават голяма част от задачите, свързани с обработката на речеви сигнали. С настоящото изследване ще покажем, че въпреки простотата си, тези модели са в състояние да обяснят и някои феномени на речеобразуването, наблюдавани при промяна на моментното психофизиологично състояние на човека. Такива са изменението на отношенията на амплитудите на формантите честоти (по-дрезгав или по-звънтящ глас), влошена разбираемост и др. За по-голяма разбираемост в по-нататъшните 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 КЪМ ПЕРИОДА НА ФОРМАНТНАТА ЧЕСТОТА

Приемаме, че моделът на вокалния тракт е от втори ред:

$$H(s) = \frac{k_1 \omega_1^2}{s^2 + \omega_1^2} \quad (7)$$

т.е. сигналът ще съдържа само един формант с кръгова честота  $\omega_1$ . При възбуждане на филтъра с последователност от правоъгълни импулси от типа:

$$e(t) = \begin{cases} 1, & (m-1)T_g \leq t < t_{open\_glottis} + (m-1)T_g \\ 0, & t_{open\_glottis} + (m-1)T_g \leq t < mT_g \end{cases} \quad (8)$$

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

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 описва както следва:

$$s_{open\_glottis}^{1F}(t) = A_{open\_glottis}^{1F} \sin(\omega_1 t + \varphi_{open\_glottis}^{1F}) \quad (9)$$

за  $t < t_{open\_glottis}$  т.е. във фазата на отворен глотис, и:

$$s_{closed\_glottis}^{1F}(t) = A_{closed\_glottis}^{1F} \sin(\omega_1 t + \varphi_{closed\_glottis}^{1F}) \quad (10)$$

за  $t \geq t_{open\_glottis}$  т.е. във фазата на затворен глотис, където:

$A_{open\_glottis}^{1F} = k_1$  е постояннотоковата съставка на сигнала при отворен глотис,

$A_{open\_glottis}^{1F} = k_1$  амплитуда на сигнала при отворен глотис

$A_{closed\_glottis}^{1F} = 2k_1 \sin\left(\frac{\omega_1 t_{open\_glottis}}{2}\right)$  амплитуда на сигнала при затворен глотис,

$\varphi_{open\_glottis}^{1F} = \frac{\pi}{2}$  е фазовите отместване на сигнала при отворен глотис

$\varphi_{closed\_glottis}^{1F} = 2\pi - \frac{\omega_1 t_{open\_glottis}}{2}$  е фазовите отместване на сигнала затворен глотис

$\omega_1 = 2\pi f_1$  е кръговата честота, съответстваща на формантната честота  $f_1$ .

Виждат се следните важни особености:

- и в двете фази - на отворен и на затворен глотис - формантната честота се определя само от параметрите на филтъра;
- амплитудата на сигнала във фазата на отворен глотис зависи само от коефициента на плоско усилване на филтъра;
- амплитудата на сигнала във фазата на затворен глотис зависи аналогично от коефициента на плоско усилване на филтъра, но и по сложна зависимост от продължителността на предхождащата фаза на отворен глотис към периода на формантната честота;
- аналогично на амплитудите стои и въпросът с фазовите отмествания на сигнала;



Практически това означава, че без да се променят параметрите на филтъра, т.е. геометрията на речевия тракт, промени на продължителността на фазата на отворен глотис могат да засилват или затихват формантните честоти на речевия сигнал. За да се види по-добре това влияние, въвеждаме безразмерен коефициент, пропорционален на времетраенето на фазата на отворен глотис към периода на формантната честота:

$$r_{\omega_1 t_{open\_glottis}} = \omega_1 t_{open\_glottis} \quad (11)$$

За отношението на амплитудите на сигнала преди и след затваряне на глотиса въвеждаме коефициента:

$$r_{A_{open\_glottis}^{1F} A_{closed\_glottis}^{1F}} = \frac{A_{open\_glottis}^{1F}}{A_{closed\_glottis}^{1F}} \quad (12)$$

Където с  $t_{closed\_glottis} = T_g - t_{open\_glottis}$  се означава времетраенето на фазата на затворения глотис в рамките на периода на основния тон.

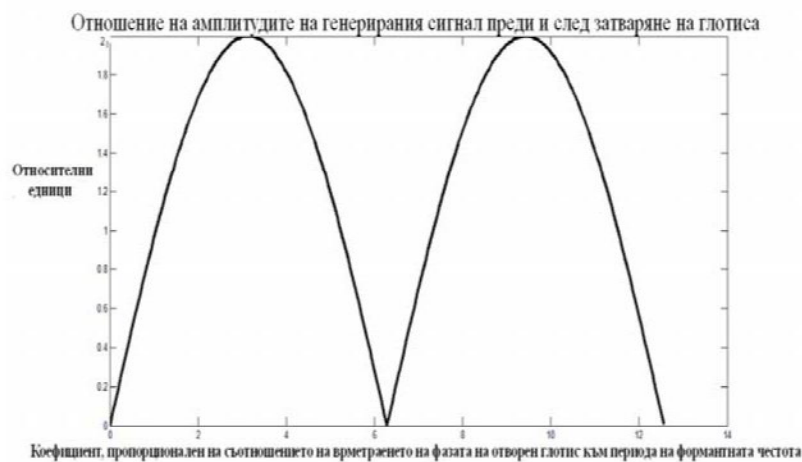
Функционалната зависимост между тези коефициенти е:

$$r_{A_{open\_glottis}^{1F} A_{closed\_glottis}^{1F}} = f\left(r_{\omega_1 t_{open\_glottis}}\right) = \left| 2 \sin\left(\frac{\omega_1 t_{open\_glottis}}{2}\right) \right| \quad (13)$$

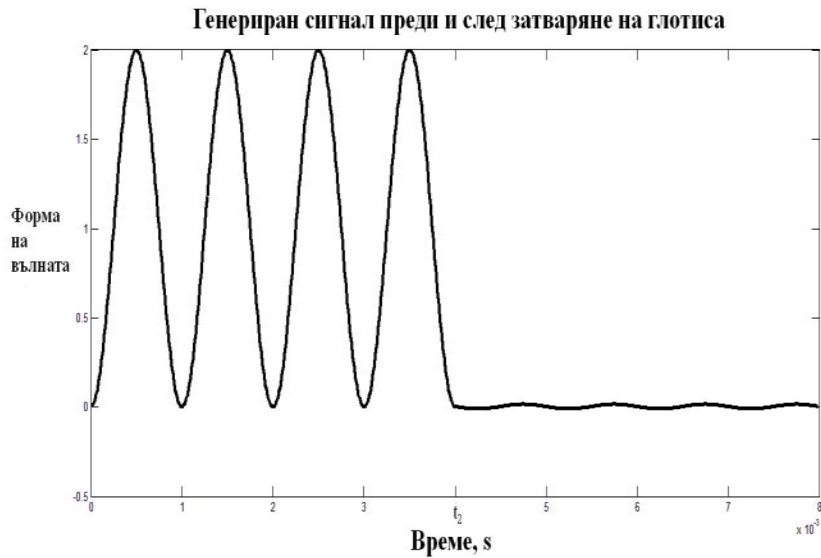
Очевидно зависимостта е периодична, като първите няколко периода са показани на фиг.2. Вижда се, че варирайки продължителността на фазата на отворен глотис и без да се променят параметрите на филтъра, т.е. геометрията на вокалния тракт, за дадена формантна амплитуда на генерирания сигнал във фазата на затворен глотис може практически да приеме която и да е стойност от нула (фиг.3) до два пъти амплитудата на отворен глотис (фиг.4).

Този ефект става още по-интересен в рамките на речеви сегмент, съдържащ няколко периода на основния тон.

Тогава, освен коефициента  $r_{\omega_1 t_{open\_glottis}} = \omega_1 t_{open\_glottis}$ , върху отношението на амплитудите ще оказва влияние и съотношението на продължителността на фазата

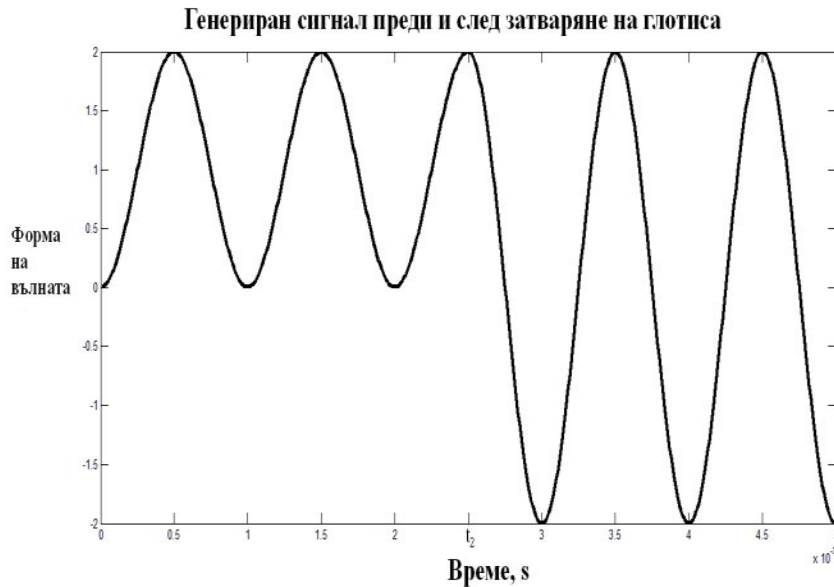


Фиг.2 Отношение на амплитудите на генерирания сигнал преди и след затваряне на глотиса като функция на коефициента  $r_{\omega_1 t_{open\_glottis}}$



Фиг.3 Генерираният сигнал преди и след затваряне на глотиса при

$$\gamma_{\omega t_{open\_glottis}} = 2n\pi, \quad n = 0, \pm 1, \pm 2, \dots \text{ и съответно } \gamma_{A_{open\_glottis}^{LF} A_{closed\_glottis}^{LF}} = 0.$$



Фиг.4 Генерираният сигнал преди и след затваряне на глотиса при

$$\gamma_{\omega t_{open\_glottis}} = n\pi, \quad n = \pm 1, \pm 3, \pm 5, \dots \text{ и съответно } \gamma_{A_{open\_glottis}^{LF} A_{closed\_glottis}^{LF}} = 2.$$

на затворен глотис към периода на основния тон:

$$k_{full\_imp} = \frac{t_{open\_glottis}}{T_g} \quad (14)$$

В този случай се поличават сравнително сложни аналитични зависимости. Примерно отношенията на амплитудите на сигнала на фазата на затворен глотис към фазата на отворен глотис за втория период на основния тон се задава с израза:

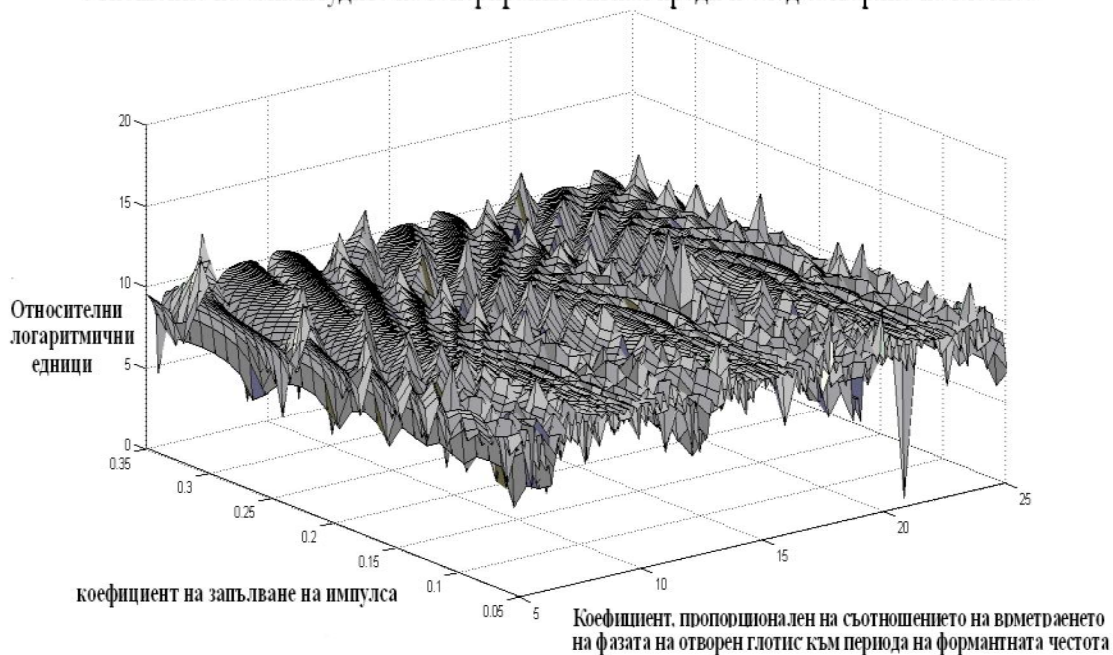
$$\frac{r_{A_{open\_glottis}^{1F}}}{A_{closed\_glottis}^{1F}} = f(r_{\omega t_{open\_glottis}}) = \left[ 4k_1 \sin\left(\frac{r_{\omega t_{open\_glottis}}}{2}\right) \cos\left(\frac{r_{\omega t_{open\_glottis}}}{2k_{full\_imp}}\right) \right]$$

$$\left[ \left( k_1 \sin(r_{\omega t_{open\_glottis}}) + k_1 \sin\left(2\pi - \frac{r_{\omega t_{open\_glottis}}}{k_{full\_imp}}\right) \right)^2 + \left( -2k_1 \sin^2\left(\frac{r_{\omega t_{open\_glottis}}}{2}\right) + k_1 \sin\left(\frac{r_{\omega t_{open\_glottis}}}{k_{full\_imp}} - \frac{\pi}{2}\right) \right)^2 \right]^{\frac{1}{2}} \quad (15)$$

и съответно с графиката от фиг.5.

Очевидно за всеки следващ период изчисляването на това отношение става все по-сложно и се налага използване на съответни числени методи.

Отношение на амплитудите на генерирания сигнал преди и след затваряне на глотиса



Фиг.5 Отношение на амплитудите на генерирания сигнал преди и след затваряне на глотиса за втория период на основния тон.

Въпреки това, може да се направи следния очевиден извод:

Променяйки съотношението на продължителността на фазите на затворен глотис към продължителността на основния тон и без да се променя геометрията на вокалния тракт, мога да се генерират речеви сегменти, при които амплитудата на даден формант за всеки следващ период на основния тон да нараства, да затихва или да не се променя съществено.

#### 4. ИЗВОДИ

Проведеното изследване показва, че моделът на *Fant* всъщност адекватно описва много повече феномени на процеса на речеобразуване, отколкото е прието от познатата теория и практика. Причина за това вероятно се крие в генезиса на самия модел. Той е създаден като модел от типа "източник-филтър", за да се



опростят процесите на анализ, на параметризация, на конкретна техническа реализация и др. Това се постига, като филтърът се разглежда практически независимо от източника на възбуждане и така за целта могат да бъдат използвани значително по-ефективни методи и средства.

## 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] Chiba, T., Kajiyama M. (1995), *The Vowel, Its Nature and Structure*, Tokyo-Kaiseikan, Tokyo, 1995
- [2] Fant G. (1990), *Acoustic Theory of Speech Production*, Mouton & Co., Hauge, 1990
- [3] Hayes M. (1999), *Schaums's Outline of Theory and Problems of Digital Signal Processing*, Singapore, McGraw-Hill, 1999
- [4] Rabiner L., Schafer R. (1992), *Digital Processing of Speech signals*, Prentice-Hall Inc., Engelwood Cliffs, New Jersey, 1992
- [5] Stevens, K., Hirano M. (1991), *Vocal Fold Physiology*, Tokyo U.P., Tokyo, 1991

**Автори:** Дамян Дамянов, асистент, катедра „Автоматизация на Непрекъснатите Производства”, Факултет Автоматика, Технически Университет - София, E-mail address: [damyan.damyandamyanov@fdiba.tu-sofia.bg](mailto:damyandamyandamyanov@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](mailto:vtg@tu-sofia.bg)

**Постъпила** на 28.04.2012

**Рецензент** проф. д-р С. Йорданова