

ПРИЛАГАНЕ НА СОФТУЕРНА СИСТЕМА ЗА АКУСТИЧЕН АНАЛИЗ ПРИ ИЗСЛЕДВАНЕ РАК НА ЛАРИНГСА

н.с.І ст. Петър Димитров Митев, ас.Азнив Карник Ангелова *
Централна Лаборатория по Биомедицинско Инженерство -БАН
*Технически Университет-София филиал Пловдив

It is well known that most laryngeal disease (esspecially cancer) cause changes in the speech. An effective and non-invasive method for laryngeal pathology evaluation and dedection is the acoustic voice analysis. In this paper a software package based on specific methods, algorithms and approaches for pathological voice analysis is proposed.

Говорните сигнали са резултат от вибрационното възбуждане на въздуха във вокалния тракт. Получаването им е акустичен и аеродинамичен феномен. Нарушаването на един от елементите на този тракт води до промени в параметрите на говорния сигнал, които са показателни за състоянието му. Един от най-важните звукообразуващи елементи е ларингса. Ларингеалните функции са разпределителна, респираторна и фонационна (генерирането на говора). Заболяванията на ларингса, респективно най-тежкото му заболяване - рака, водят до нарушаване нормалната фонация на субекта, като предизвикват промяна на говора.

Един ефективен и неинвазивен метод за оценка ларингеалната патология и откриването ѝ, е акустичния анализ на говорните сигнали. Доказано е [1], че при повечето пациенти, ларингеалната патология се характеризира с: увеличена степен на шумовете, широки граници на промяна на периода на основния тон и амплитудата на пиковете на основния тон, прекъсвания в генерирането на основния тон, наличие на субхармонични компоненти в спектъра на говорния сигнал, деформирани кривата на основния тон, наличие на високочестотни шумови компоненти.

Софтуерната система, която е представена в този доклад, дава възможност за акустичен анализ на говорните сигнали във времевата, спектралната и кепстралната област и има следните предимства: анализа е напълно неинвазивен, системата е изградена на базата на персонален компютър със стандартен допълнителен хардуер - стандартна звукова карта, софтуера е графически описан и лесно използваем и може да се инсталира на персонален компютър за извършване на периодичен скрининг, системата е електрически защитена тъй като няма галваническа връзка между пациента и персоналният компютър.

Анализ във времевата област:

Системата е специализирана за патологични гласове и във връзка с това са използвани следните специфични методи, алгоритми и приложения:

1. Дедектор с повишена точност за отделяне на периодичните от непериодичните сегменти на говорния сигнал с използване метода на невронните мрежи(PAS): Квантования сигнал се дели на сегменти с дължина T_0 , където T_0 представлява периода на основния тон - T_0 е оптималната дължина на сегмента [2]. Коректното определяне на T_0 е изключително важно за определяне останалите параметри на говорния сигнал. За елиминирание грешките от неточностите при T_0 се използва дедектора с повишена точност [3].

Откриване пиковите на основния тон и тяхното разположение със средствата на метода на геометричния център на масата : При нормалните гласове съществува един най-голям максимум (пик), а при патологичните гласове повече от един съответстващи пикове основната честота на тона.. За откриването на истинския пик на основния тон(PPP) се използва метода на геометричния център на масата(GCM). Процедурата от Фигура1 се използва за изчисляване на PPP .

Откриване на четирите най-големи максимума- $A_{\max}(i)$ при $i=1, \dots, I=4$, и техните минимуми(най-малката стойност на разстояние по-малко от $0.2 T_0$ наляво от максимума)

Изчисляване на двете стойности:

$$P+ = k \max_{i=1}^I \frac{A_{\max}(i)}{I} \quad (1) \quad P- = k \min_{i=1}^I \frac{A_{\min}(i)}{I} \quad (2)$$

$k_{\max} = 0.7$ и $k_{\min} = 0.6$ (експериментално установени).

Центриране изразвянето на сигнала , използвайки $P+$, $P-$ [4]

Откриване на отделните периоди $T_0(j)$,за $j=1, \dots, J$ и пулсациите на основния тон- $A_{\max}(i)$ (за $i=1, \dots, I$) [4]

Изчисляване на : $TRMAX(i) = 0,9 A_{\max}(i)$ (3)

Проверка за наличие на пикове по-големи от $TRMAX(i)$ и на разстояние по-малко от $0,2T_0$ в ляво и в дясно от $A_{\max}(i)$

Изчислява се

$$TRgm(i) = b A_{\max}(i) \quad (4)$$

$A_{\max}(i)$ - амплитудата на i -ия максимум на основния тон , $b=0,4$

Определяне броя ($L1$) на дискретите наляво от $A_{\max}(i)$ с амплитуди по-големи от $TRgm(i)$ и броя($L2$) на дискретите надясно от $A_{\max}(i)$ с амплитуди по-големи от $TRgm(i)$.

Изчисляване на PGCM(за i тия максимум на основния тон:

$$PGCM(i) = i + \frac{\sum_{j=-L1}^{L2} x(i+j)j}{\sum_{j=-L1}^{L2} x(i+j)} \quad (5) \quad A_{\max}(i) = x(PGCM(i)) \quad (6)$$

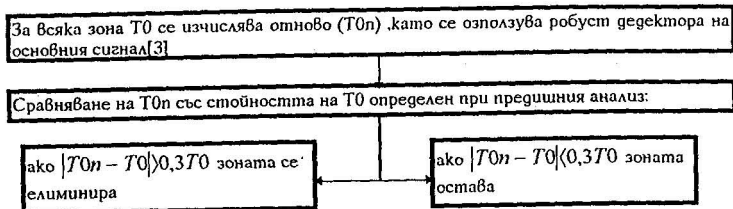
i - индекс на дискрета на i -я максимум на пиковите на основния тон
 $x(i+j)$ дискретите на сигнала

Фигура 1

Формиране на стабилни зони , със стабилен период на основния тон и T_0 приблизително константа:Алгоритъма ,който се използва, за установяване периода на основния тон T_0 [3] практически отстранява почти всички отклоняващи се стойности на T_0 . Изчислява се средната стойност на

$T_0(T_{0m})$ за всички сегментите и тя се използва за изчисляване размера на прозореца при сегментацията на сигнала.

Елиминирането на "лошите" стойности на периода на основния тон чрез двустъпалното му анализиране: Процедурата е дадена на Фигура 2



Фигура 2

Със цел да се направи количествена оценка на стабилността на генерациите на основния тон се изчисляват следните нови параметри:

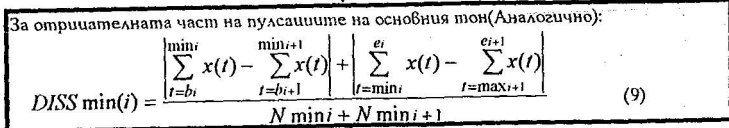
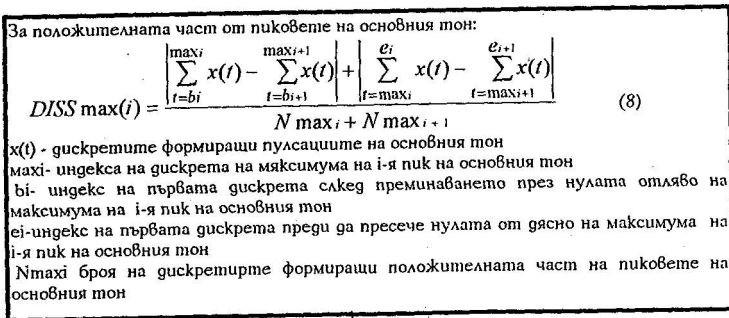
Степен на стабилност на генериране на основния тон (STAB). Стабилността на основния тон се изчислява с формулата:

$$STAB = Nstab / N \quad (7)$$

$Nstab$ - броят на сегментите в цялата стабилна зона

N - броят на сегментите във всички периоди.

Степен на различност в кривите на пулсациите на основния тон (DISS). Изчисляването на тази величина става по процедурата от Фигура 3:



А

A

↓

Изчисляване средната стойност на несъответствията на пулсациите на основния тон:

$$DISS(h) = \frac{1}{I} \sum_{i=1}^{I-1} [DISS \max(i) + DISS \min(i)] / 2 \quad (10)$$

I - броят на пиковите на основния тон в зоната
 DISS(h)-степента на деформиране за h зоната.

Фигура 3

Стабилността на T0 генерациите се анализира с използването на две добре познати величини: коефициент на смущение на периода на основния тон(PPQ) и коефициент на смущение на амплитудата на основния тон (APQ). За изчисляването на тези величини се използва допълнителната величината-относителна средна девиация(RAP) [5].RAP се изчислява с формулата:

$$RAP = \frac{1}{N-2} \sum_{i=2}^{N-2} \frac{[P(i-1) + P(i) + P(i+1)] / 3 - P(i)}{\frac{1}{N} \sum_{i=1}^N P(i)} \quad (11)$$

RAP - PPQ или APQ

P(i) е стойността на i-я период или амплитуда на основния тон

N е броят на всички периоди или амплитуди при фонацията.

Степен на невокализираност(DUV) - този параметър дава съотношението на беззвучните към пълния брой звучни сегменти по време на фонацията. Изчисляването става с формулата :

$$Duv = Nuv / Ntot, \quad (12)$$

Ntot - пълния брой звучни сегменти

Nuv - броят на беззвучните сегменти

Анализ в спектралната област:

Формиране на стабилни сегменти: За изчисляване динамиката на спектралните параметри стабилната зона се дели на S на брой стабилни сегмента. Само зони съдържащи повече от 2048 точки се включват за изследване.

Спектрален анализ-Използува се 2048 точкова бърза Фурие трансформация.

Съотношение ниска / високо честотна енергия(RLH)

$$RLH(s) = Elow(s) / Ehigh(s) \quad (13)$$

Elow(s) - енергията при честота 0-1500 Hz

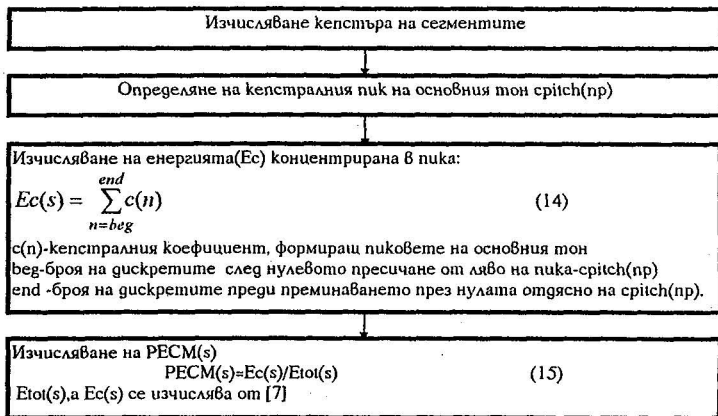
Ehigh(s) е енергията при честота 1500-4000 Hz.

Анализ в кепстралната област:

Кепстъра на патологичните гласове се характеризира с изкривяване на кривата на пулсациите на основния тон и променени амплитуди на тези пулсации[6,2]. Кепстъра се получава по процедурата: изчисляване на спектъра с използване на FFT, логаритмуване на изчисления спектър, прилагане на

обратното FFT върху спектъра. Един от важните параметри, който се използва е:

Съотношението(Енергията концентрирана в кепстралния пик на основния тон към пълната кепстрална енергия(РЕСМ)-Фигура 4



Фигура 4

Други специфични изчисления при патологичните гласове:

Изчисляване на количеството шумови компоненти(QNC)-Наличието на шумови компоненти в гласа се провокират от ларингеалната дисфункция.. Количеството им се изчислява в спектралната като съотношение шум към хармоници(DH) [8], и времевата област. а във времевата област се изчислява като хармоници към шум(NHR)[9]

Глобален анализ-Глобалните вълни се получават от синхронизирано с основния тон итеративно адаптивно инверсно филтриране[10]

На базата на този софтуерен продукт в продължение на период от една година във специализирания кабинет на Фонеатричната клиника на Медицинска Академия -София са записани гласовете на 50 хора с нормална фонация и 50 пациента с проблеми при фонацията. Всички изговарят гласната "а" на всеки 4 месеца. Сигнала се квантова с честота 20480 Hz и точност 12 бита като се използва "DSP Sonagraph 5500". Анализа показва твърде значителни промени в параметрите на PPQ,DH,DISS,DUV и по-малки по отношение на APQ,HNR и .Резултатите от изследванията на тези параметри за пациент (21-104-мъж) с патология на ларинкса са дадени в Таблица1(подчертаните стойности са извън нормата):

	<i>T0</i> [ms]	<i>PPQ</i> [%]	<i>APQ</i> [%]	<i>DUV</i> [%]	<i>HNR</i> [dB]	<i>DH</i> [abs]	<i>RLH</i> [dB]	<i>PECM</i> [abs]
01.97г.	4,772	2,16	3,89	0	14,04	0,398	10,328	0,490
04.97г.	3,975	1,35	3,52	0	2,67	0,4623	11,775	0,405
08.97г.	3,172	1,46	4,393	0	0,00	0,4612	5,171	0,558
12.97г.	3,593	1,44	4,77	0	12,85	0,8285	5,149	0,193
нормални средни стойности(мъже)		0,28	0,92	99,8	14,52	0,28	1,03	0,372
патологични средни стойности(мъже)		0,81	1,4	94,0	3,75	0,35	0,92	0,302

Таблица I

Литература:

1. Kasuya H., Ogana S., Kikuchi J., Ebihara S.: An acoustic analysis of pathological voice and its applications to the evaluation of laryngeal pathology. *Speech Commun.* vol.5, pp171-181, 1986
2. Bojanov B. Analysis of atiological voice. Report under contract ERB-CIPA-CT-92-0170 with the European Community, 1993
3. Bojanov B., Hadjitodorov S., Chollet G.: Robust periodicity/aperiodicity detector. *Ann of Bulgarian Academy of Sciences* vol.49 N1, 1996
4. Hess W.: Pitch determination of speech signals. Springer Verlag, N.Y., 1983
5. Koike Y.: Application of some acoustic measures for the evaluation of laryngeal disfunction. *Studia Phonologica* vol.5 pp.717-723, 1973
6. Bojanov B., Chollet G.: Pathological voice analysis using cepstra, bispectra and group delay functions. In Proceedings of International Conference on Spoken Language Processing, Alberta, Canada, vol.2, 1039-1042, 12-16, October, 1992
7. Dejonkere P.: Bruit de turbulencwe et aperiodicity dans la voix pathologique. Une approude multifonctionnelle. *Revue de Laryngologie* vol.11, pp 353-357, 1990
8. Junik M. and Bojanov B.: Method for evaluation of the noise-to-harmonic component ratios in pathological and normal voices. *Acustica*, vol.70, pp 89-091, 1990
9. Jumoto E.: The quantatative evaluation of hoarseness. A new harmonics to noise ratio method. *Archives of Othoralynology*-vol109 pp48-52, 1983
10. Alku P. and Vilkmán E.: Amplitude domain quotient for characterization of the global volume velocity waveforms estimated by IF. *Speech Comm.* v.18, 131-138, 1996