خوارزمية تحديد المقاطع النشطة والصامتة في الإشارة الكلامية

الدكتور المهندس: حسان محمد أحمد

مدرس في قسم هندسة الاتصالات والشبكات – كلية هندسة الحاسوب والمعلوماتية والاتصالات – الجامعة السورية الخاصة

Hassan.Ahmad@spu.edu.sy

الملخص

يهدف هذا البحث إلى دراسة وتطوير خوارزمية تحديد المقاطع النشطة والصامتة (active segments & pauses) في الإشارة الكلامية.

تعتمد الخوارزمية المقترحة على استخدام نظرية الإدراك النشط (active perception) المكيّفة لأجل تحليل الإشارات الكلامية، وكذلك على خوارزميات تقطيع الإشارة الكلامية (speech signal segmentation).

تؤكد نتائج التجارب التي أجريت باستخدام الخوارزمية المقترحة على إمكانية استخدامها لحل المسألة المحددة والمفروضة في مجال معالجة الإشارة الكلامية (نظم التعرّف على المتحدث باستخدام بصمة صوته).

الكلمات المفتاحية: معالجة الإشارة الرقمية، نظرية الإدراك النشط، تحليل الإشارة الكلامية، تقطيع الإشارة الكلامية، المقاطع النشطة، المقاطع الصامتة.

Algorithm of isolating the pauses in a speech signal

Dr. Hassan M. Ahmad, Syrian Private University
Hassan.Ahmad@spu.edu.sy

Abstract

The article considers an algorithm of isolating the pauses in a speech signal.
The developed algorithm is based on the use of the theory of active perception adapted for the analysis of speech signals and on speech signal segmentation.
Results of the experiments confirm the possibility of the offered algorithm application for solving the specific problem in speech processing area (Speech Recognition System).

Keywords: digital signal processing, theory of active perception, speech signal analysis, speech signal segmentation, active segments, pauses isolation in a speech signal.

1. مقدمة

تعدّ مسألة التقطيع (segmentation) إحدى أهم مسائل المعالجة الرقمية وتحليل الإشارات الكلامية بهدف تحديد المقاطع النشطة والصامتة.

في هذا العمل، نعرّف المقطع الصامت (pause) على أنه المكان في الإشارة الكلامية الذي لايوجد فيه نشاط كلامي، أي خال من الكلام.

يعدُّ التقطيع واحداً من أول مراحل حل المسائل المتعلقة بمعالجة الإشارة الرقمية الكلامية، على سبيل المثال: مسائل تحسين الإشارة الكلامية، مسائل التعرف على الكلام، وغيرها... يستخدم تقطيع الإشارة في نظم الاتصالات لأجل تخفيض حجم البيانات المرسلة على حساب حذف المقاطع الصامتة.

عادةً يتحدد الطول الزمني للمقطع الصامت من خلال السبب المؤدي لتشكيله. تتشكل المقاطع الصامتة أثناء الكلام عندما تكون أعضاء النطق (articulation organs) في وضعية مغلقة، وهذا يكون مرتبطاً بلفظ الأحرف الساكنة. يشكل طول مثل هذه المقاطع الصامتة $0.1s$ [24]. أثناء عملية القراءة، يكون طول المقطع اللفظي الصامت (linguistic pause) عند حدود سلسلة من الكلمات المركبة (syntagmatic) أو عند نهاية مقطع صرفي (morpheme) لايتجاوز $0.75s$، بينما يتغير بين الجُمَل من 0.5 إلى $1.5s$ [8−12−28].

لذلك سنستخدم في هذا العمل احتمالية أن طول المقطع الصامت يمكن أن يتغير مابين 0.1 إلى $1.5s$.

تعد مسألة وجود الضجيج (noise) في الإشارة الكلامية الأولية إحدى أهم المشاكل التي تظهر أثناء عملية التقطيع. بناءً على ذلك فإنه من الضروري أن تكون خوارزمية التقطيع ذات استقرارية جيدة (stability) تجاه الأنواع المختلفة من الضجيج التي يمكن أن تصادف في الإشارة الكلامية.

تجدر الإشارة إلى وجود آليات إدراك ذات مناعة ضد الضجيج (noise−immunity) في النظام السمعي عند الإنسان[23]، ومنها:

1) وجود التغذية المرتدة الصادرة (efferent feedback) يخفض حساسية قوقعة الأذن (cochlea) عند وجود الضجيج، ويخفض أيضاً من خطر زيادة الحمل (overload)، وبذلك تحمي القوقعة من تأثير الصوت العالي [25]؛
2) وجود التفاعل السمعي الثنائي (binaural interaction) للقناتين اليمينية واليسارية يسمح بزيادة وضوح الكلام؛
3) وجود العضلة الرِّكابِيَّة (stapedius muscle) للأذن الوسطى يسمح بزيادة صلابة سلسلة العظام السمعية، مما يؤدي إلى انخفاض الطاقة الناجمة عنهم، و يسمح بتعويض تشويشات الكثافة العالية.

2. هدف البحث

الهدف الرئيس من هذا البحث هو تطوير خوارزمية لتحديد المقاطع النشطة والصامتة في الإشارة الكلامية بالاعتماد على نظرية الإدراك النشط.

3. استعراض الخوارزميات المتوفرة

نستعرض بعض الخوارزميات المتوفرة لتحديد المقاطع النشطة والصامتة في الإشارة الكلامية.

يقدم الباحث Omid Ghahabi وزملاؤه في عملهم [15] تقنية تحديد النشاط الصوتي المنيع لأجل نظم التعرف التلقائي على الكلام في الزمن

الحقيقي والمبنية على أساس إحصائيات وخوارزمية Baum–Welch .

يقدم الباحث Weiqing في عمله [21] تقنية تحديد النشاط الصوتي في الزمن الحقيقي باستخدام مرشح Gammatone.

يقدم الباحث Alimuradov في أعماله [1–2]، الأول: خوارزمية الضجيج المنيع لتقطيع الكلام/الصمت في نظم التشخيص، حيث يستخدم الباحث في خوارزميته المقترحة طرائق المعالجة التكيفية للإشارات غير المستقرة (non–stationary) والمعالجة الإحصائية للبيانات (statistical data processing) وكذلك طريقة التفاضل باستخدام مفاهيم التوزيع الطبيعي. وفي الثاني: خوارزمية تحديد الصمت والكلام في الإشارةالكلامية مستخدماً الطريقة التكيفية للتحليل المتكامل لوظائف الحالة الداخلية للمتكلم.

يقدم الباحث Ortiz في عمله [5] خوارزمية بسيطة وفعالة لتحديد النشاط الصوتي مستخدماً تحويل هيلبرت (Hilbert transform) والعتبة الديناميكية لأمراض ومشاكل الكلام.

يقدم الباحث Beritelli في عمله [3] طريقة لتقطيع الإشارات مبنية على أساس نظرية التعرف على الكلام ونظرية المنطق المبهم (fuzzy logic). يمكن استخدام هذه الطريقة لمعالجة الإشارات المضجة.

يقدم الباحث McKinley في عمله [13] خوارزمية تحديد النشاط الصوتي على أساس الحد الأعظمي للمعلومات الاستدلالية (posteriori information) ومعيار (criterion) نيمان–بيرسون (Neyman–Pearson).

يقدم الباحثان Sohn و Cho في أعمالهما [4–19] خوارزمية تحديد المقاطع النشطة والصامتة المبنية على فرضية أنّ تحويل فوريية المتقطع (discrete Fourier transform) للكلام والضجيج يعدّ متغيرات عشوائية غوصية (Gaussian) مستقلة ومتقاربة.

يقدم الباحث Gazor في عمله [7] خوارزمية تحديد نشاط الكلام المبنية على استخدام تحويل جيب التمام المتقطع (discrete cosine transform)، حيث تم تصميم الخوارزمية مع الأخذ بعين الاعتبار حقيقة أنّ قانون توزيع الكلام والضجيج يتوافق مع نماذج غاوص ولابلاس (Gaussian and Laplace models).

يقدم الباحث Sheikhzadeh في عمله [18] خوارزمية تقطيع الإشارة على أساس حساب دالة الارتباط الذاتي للإشارة. تم تطوير هذه الخوارزمية لنظم الزمن الحقيقي وقد تم تنفيذها على منصة معالجة الإشارة الرقمية (DSP–platform) لأجل تحسين الكلام في الأجهزة السمعية.

يقدم الباحث Rezayee في عمله [17] خوارزمية تحديد المقاطع الصامتة المبنية على أساس تحويل كارانن لوف التكيفي (Karhunen–Loève adaptive transformation).

يقدم الباحثان Everteletskaya و Vanoop في أعمالهما [6–20] طرائق و تقنيات تحديد النشاط الصوتي لأجل تحسين الكلام مستخدمين طريقة تحديد النشاط الصوتي المنيع للضجيج و تقنية التحسين الأمثل الهجين (Hybrid Optimization).

نرى من الضروري الإشارة إلى أن الطرق والخوارزميات المتوفرة لتحديد المقاطع الصامتة في الإشارة الكلامية تسعى دائماً لتحسين واحداً من البارامترات الآتية: دقة تحديد المقطع الصامت، التأخر الزمني (time delay) والتعقيد الحسابي (computational complexity) [9–10–14–15–21–22].

لأجل تقييم نوعية الإشارة الكلامية نستخدم نسبة الإشارة إلى الضجيج (signal–to–noise ratio, SNR) [16]:

$$SNR = 10\log_{10}\left(\sum_{n=1}^{N_s} s^2(n) \Big/ \sum_{n=1}^{N_v} v^2(n)\right) = 10\log_{10}\frac{P_s}{P_v} \quad (1)$$

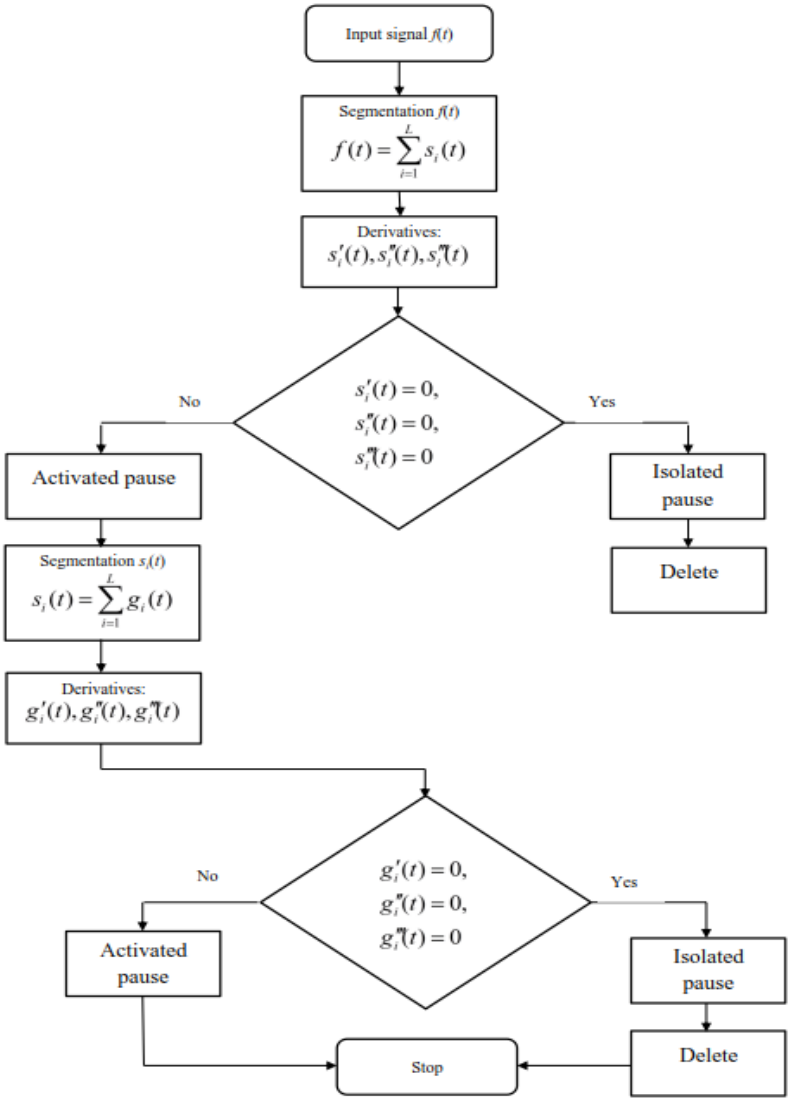
حيث:

$s(n)$ – الإشارة الكلامية؛ $v(n)$ – إشارة الضجيج؛ N_s – عدد العينات (samples) في الإشارة الكلامية؛ N_v – عدد العينات في إشارة الضجيج؛ P_s – استطاعة الإشارة الكلامية و P_v – استطاعة إشارة الضجيج.

4. الخوارزميات المقترحة

4.1 خوارزمية تحديد المقاطع الصامتة في الإشارة غير المضجة.

ليكن $f(t)$ الإشارة الكلامية الأولية المراد تحليلها. تتلخص مراحل الخوارزمية المقترحة في المخطط الانسيابي المبين في الشكل (1).



الشكل (1): المخطط الانسيابي لخوارزمية تحديد المقاطع الصامتة في الإشارة غير المضجة

توصيف المخطط الانسيابي المبين في الشكل (1):

تتلخص الخوارزمية المقترحة في التحليل التسلسلي الآتي للإشارة:

1) تقسيم الإشارة الأولية $f(t)$ إلى مقاطع بطول L عينة؛
2) حساب الاشتقاق (derivation) [11] الأول والثاني والثالث لكل مقطع؛
3) إذا كانت اشتقاقات المقطع ذي التسلسل i مساوية للصفر، فمن الضروري تسجيل أنّ هذا المقطع هو مقطع صامت، وبالتالي حذفه (عذله) من عملية المعالجة اللاحقة.
4) إذا كانت اشتقاقات المقطع ذي التسلسل i غير مساوية للصفر، فإننا نستمر في تقسيم المقطع i الذي يجرى تحليله إلى مقاطع جزئية أخرى، ومن ثم تطبيق الخطوات السابقة 2 و 3 على كلٍ من المقاطع الجزئية الناتجة.

تنهي الخوارزمية عملها عندما يتم تصنيف كل مقطع من مقاطع الإشارة أو عند الوصول إلى مستوى معين من الدقة. إذا توصلنا إلى مستوى معين من الدقة، وبقي بعض المقاطع التي لم يتم تصنيفها، والتي اشتقاقاتها لاتساوي الصفر، فإنّه يتم تسجيل مثل هذه المقاطع على أنّها مقاطع كلامية نشطة (غير صامتة).

إذا أخذنا بعين الاعتبار أن الطول الأصغري للمقطع الصامت هو 0.1s فإنه بإمكاننا حساب الطول الأصغري للمقطع من الإشارة الذي يجري تحليله من خلال تردد أخذ العينات (sampling frequency) المعلوم مسبقاً:

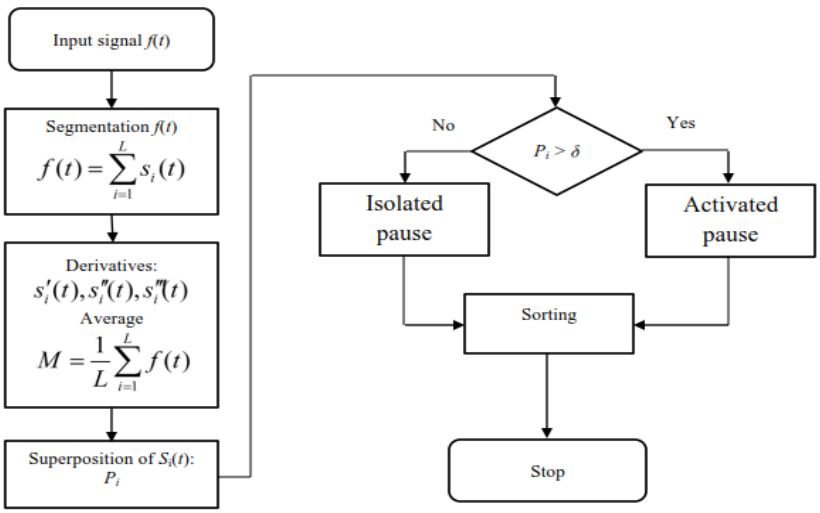
$$L = 0.1F \quad (2)$$

حيث: F – تردد أخذ العينات.

4.2 خوارزمية تحديد المقاطع الصامتة في الإشارة المضجة

تسمح الخوارزمية التي جرى توصيفها سابقاً بتقطيع الإشارة غير المضجة. لكن في الظروف الحقيقية لايمكن عملياً أن نجد مثل هذه الإشارات بسبب وجود التشوهات في خطوط الاتصالات والأجهزة المستخدمة و وجود الضجيج من البيئة المحيطة. بالتالي، فإن خوارزمية تحديد وعزل المقاطع الصامتة لابد أن تأخذ بعين الاعتبار وجود الضجيج في الإشارة.

تتلخص الخوارزمية المقترحة لتحديد وعزل المقاطع الصامتة في الإشارة المضجة في المخطط الانسيابي المبين في الشكل (2).



الشكل (2): المخطط الانسيابي لخوارزمية تحديد المقاطع الصامتة في الإشارة المضجة

توصيف المخطط الانسيابي المبين في الشكل (2):

1) تقسيم الإشارة الأولية $f(t)$ إلى مقاطع بطول L عينة؛
2) حساب الاشتقاق (derivation) الأول والثاني والثالث والمتوسط (average) $M=\frac{1}{L}\sum_{i=1}^{L} f(t)$ لكل مقطع؛
3) حساب تراكب (superposition) كل قيم البرامترات الناتجة لكل مقطع، فإذا كانت نتيجة حساب التراكب للمقطع ذي التسلسل i أكبر من قيمة عتبة معينة، فإنه يمكن الافتراض بأنّ هذا المقطع يحتوي على نشاط كلامي.
4) تحديد مقاطع الإشارة التي يتركز فيها النشاط الكلامي؛
5) تنفيذ آلية تمدد/ضغط (expansion/ compression) المناطق التي يتركز فيها النشاط الكلامي بهدف تحديد حدود مناطق الإشارة التي تنتمي إلى مناطق النشاط الكلامي (تشبه هذه الآلية خوارزمية تمدد/ ضغط مناطق المجال البصري [29]).

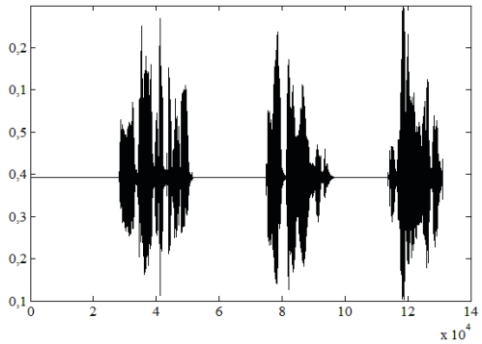
تنتهي الخوارزمية عملها عندما يتم تصنيف كل مقطع من مقاطع الإشارة.

5. التجربة الحسابية

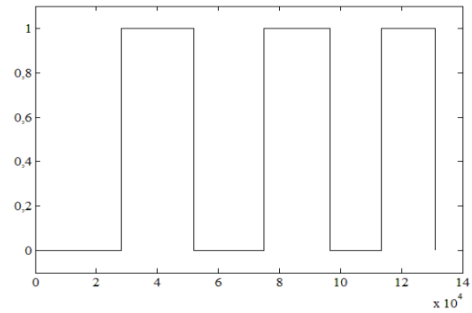
أجريت التجارب الحسابية وتسجيل الإشارات باستخدام بيئة ماتلاب (Matlab)، وباستخدام حاسب بمعالج Pentium Dual Core T4200 (2 GHz) وبذاكرة 4 GB.

يبين الشكل (3) الإشارة الكلامية المعدة للاختبار: تردد أخذ العينات 16kHz و عمق الترميز (coding depth) 16bit وبطول زمني 8.2s.

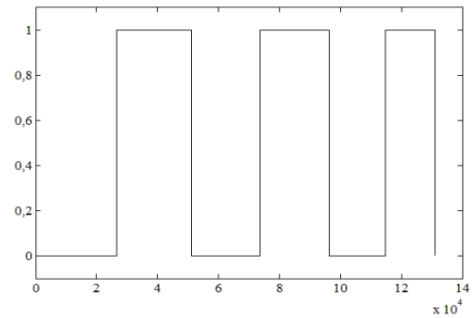
يبين الشكل (4) الإشارة الكلامية المقطعة يدوياً. يبلغ الحجم الأصغري للمقطع 1024 عينة (64ms) وفقاً لبارامترات الإشارة المشار إليها سابقاً. عند حساب طول المقطع وفقاً للعلاقة (4) (انظر الشكل (5)) يتم اختيار الطول الناتج بمثابة عدداً، يكون أساً للعدد 2. بالتالي نختار الطول الأصغري للمقطع $L=128$ عينة (8ms). تجدر الإشارة إلى أنّ النظام السمعي يُقيّم التغيرات التي تحصل خلال فترة زمنية أقل من 1ms في النبضة الصوتية [29].



الشكل (3): الإشارة الكلامية الأولية

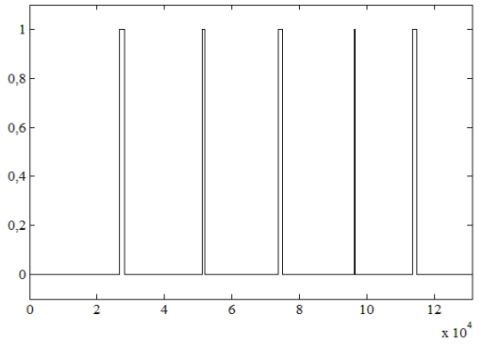


الشكل (4): التقطيع اليدوي للإشارة الكلامية



الشكل (5): التقطيع الآلي للإشارة الكلامية

بينت التجارب الاختبارية التي أجريت على الإشارة الكلامية الأولية باستخدام الخوارزمية المقترحة (الشكل (6)) أن الخطأ من النوع الأول (خطأ تحديد المقطع الصامت) قد بلغ 0.02، بينما بلغ الخطأ من النوع الثاني (تجاوز المقطع الصامت) 0.017 و قد بلغ زمن معالجة الإشارة الأولية 0.18s.



الشكل (6): الفرق بين التقطيع اليدوي والآلي

أثناء اختبار الخوارزمية المقترحة على الإشارة الكلامية المضجحة تم استخدام الأنواع الآتية من الضجيج الجمعي:

1) النظرية (theoretical):
   - ضجيج جمعي طبيعي (غوصي) (Additive normal noise (Gaussian))؛
   - ضجيج جمعي منتظم (Additive uniform noise)؛
2) الحقيقية (real):
   - ضجيج طريق سريع (اتوستراد)؛
   - ضجيج الشارع؛
   - ضجيج محرك السيارة.

يعدّ ضجيج الطريق السريع (اتوستراد) شبيهاً بالضجيج الطبيعي. يتجلى الفرق بينهما في حقيقة أنّ ضجيج الطريق السريع يحتوي على مقاطع تزداد فيها الكثافة (intensity) عند اقتراب السيارة من أدوات تسجيل الصوت.

يبين الجدول (1) نتائج التجارب الحسابية التي أجريت على الإشارة الكلامية المضجحة بأنواع مختلفة من الضجيج. يشير العمود err1 إلى الخطأ من النوع الأول والذي يبين النسبة المئوية لكمية المقاطع الصامتة التي جرى تحديدها بشكل كاذب (false pause)، بينما يشير العمود err2 إلى الخطأ من النوع الثاني الذي يبين النسبة المئوية لكمية المقاطع الصامتة التي تم تجاوزها (skipped pauses).

يبين الجدول (2) نتائج تقييم جذر مربع متوسط انحراف الضجيج (Root-Mean-Square Deviation, RMSD).

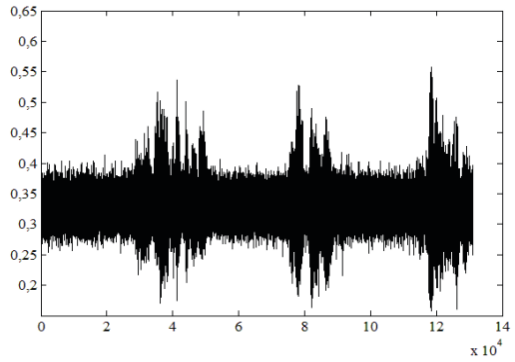
الجدول (1): نتائج عمل الخوارزمية

SNR	نوع الضجيج									
	ضجيج طبيعي		ضجيج منتظم		ضجيج اتوستراد		ضجيج الشارع		ضجيج محرك سيارة	
	err1	err2	err1	err2	err1	err2	err1	err2	err1	err2
30 dB	0	6	0	6	5	5.6	0	5.8	0	5
20 dB	0	6	0	6.7	4.3	5.9	0	4.2	0	6
10 dB	0	7.1	10	3.8	3.7	3.9	5.3	3.3	1.5	3.4
5 dB	1.5	7.8	32.4	1.9	3.8	4.4	20.1	1.5	14.8	3.5
0 dB	31.4	4	51.8	0	4	2.4	20.7	3.6	40.1	3.7

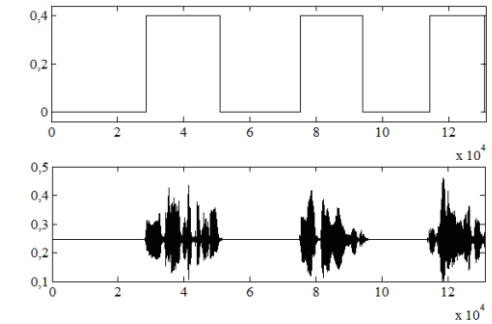
الجدول (2): خصائص الضجيج (RMSD للضجيج)

SNR	نوع الضجيج				
	ضجيج طبيعي	ضجيج منتظم	ضجيج اتوستراد	ضجيج الشارع	ضجيج محرك سيارة
30 dB	0.00171	0.00405	0.00091	0.00083	0.00089
20 dB	0.00559	0.01327	0.00454	0.00417	0.00445
10 dB	0.02021	0.04034	0.01362	0.00835	0.01423
5 dB	0.03495	0.05762	0.02088	0.01669	0.02313
0 dB	0.05013	0.11537	0.84024	0.03004	0.03503

تبين الأشكال (7) و (8) و (9) إشارة كلامية أولية ونتيجة تقطيعها.



(a)

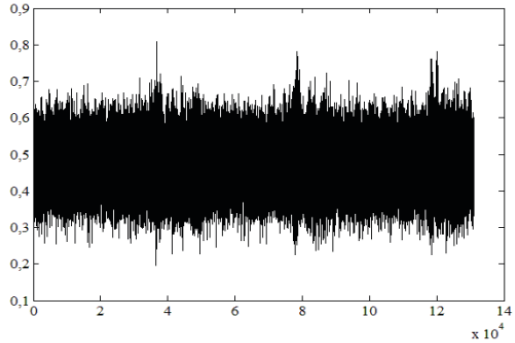


(b)

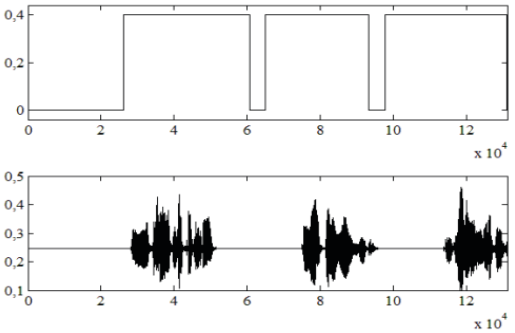
الشكل (7): إشارة مضجة بضجيج طبيعي

($SNR = 10\text{dB}$)

(a) – الإشارة الأولية، (b) – نتيجة التقطيع



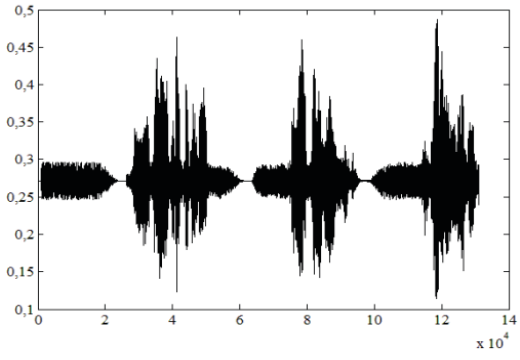
(a)



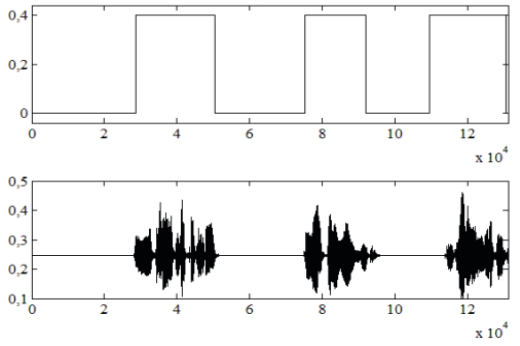
الشكل (8): إشارة مضجة بضجيج طبيعي

($SNR = 0\text{dB}$)

(a) – الإشارة الأولية، (b) – نتيجة التقطيع



(a)



(b)

الشكل (9): إشارة مضجة بضجيج محرك سيارة

($SNR = 20\text{dB}$)

(a) – الإشارة الأولية، (b) – نتيجة التقطيع

5. الاستنتاجات

تتجلى نتائج دراسة الخوارزمية المقترحة المناعية للضجيج في الآتي:

1) تنخفض دقة التقطيع عند زيادة الطول الزمني للإشارة التي تتم معالجتها. احد الحلول الممكنة لهذه المشكلة هو المعالجة المحددة لمقاطع الإشارة بطول 2–3 ثانية.
2) إذا قارنا دقة تحديد حدود المقاطع S_1 و S_2 (حيث: A_1 المطال الأعظمي للمقطع S_1 و A_2 المطال الأعظمي للمقطع S_2؛ $A_1 \ll A_2$) فإن دقة تحديد حدود المقطع S_1 تكون أقل من دقة تحديد حدود المقطع S_2، وذلك بسبب اختلاف مستوى تأثير الضجيج المرفق للإشارة على مقاطعها المختلفة المطالات.

3) إذا قارنا دقة تحديد حدود المقاطع S_1 و S_2 (حيث: L_1 طول المقطع S_1 و L_2 طول المقطع S_2؛ $L_1 \ll L_2$) فإن دقة تحديد حدود المقطع S_1 تكون أقل من دقة تحديد حدود المقطع S_2.

تجدر الإشارة إلى أنّ تقييم الخوارزمية المقترحة وفقاً للبارامتر "دقة التقطيع" ليس أقل أهمية من الخوارزميات المقترحة لتقطيع الإشارة الكلامية [16]، وتظهر في بعض الحالات نتائج أفضل.

6. الخلاصة

تم في هذا العمل دراسة وتطوير خوارزمية تقطيع الإشارة الكلامية إلى مقاطع نشطة الكلام وصامتة. نقدم في هذا العمل خوارزميتان: الأولى مخصصة لتقطيع الإشارات المضجة، والثانية لتقطيع الإشارات غير المضجة. عند تصميم الخوارزميتين المذكورتين تم الأخذ بعين الاعتبار طرائق مبنية على أساس الآليات المماثلة التي تعمل في نظام الإدراك البصري.

يمكننا أيضاً تمييز الخوارزمية المقترحة مقارنة مع الخوارزميات الأخرى المتوفرة بالنقاط الإيجابية الآتية:

- انخفاض التعقيدات الحسابية؛
- عدم وجود بارامترات تحتاج إلى معايرة؛
- سهولة التنفيذ والتطبيق؛
- دقة تحديد حدود المقاطع النشطة والصامتة.

المراجع

1. ALIMURADOV. A, K, TYCHKOV. A, YU, CHURKOV. P, P, KVITKA YU, S, ZARETSKY. A, P, VISHNEVSKAYA. G, V, **2016** "Noise-Robust Algorithm for "Speech/Pause" Segmentation in Diagnostic Systems of Psychogenic States", 2016 International Conference on Engineering and Telecommunication (EnT), pp. 3-6, Nov. 29-30.
2. ALIMURADOV. A, K, TYCHKOV. A, YU, **2017** Speech/pause detection algorithm based on the adaptive method of complementary decomposition and energy assessment of intrinsic mode functions. Soft Computing and Measurements (SCM), XX IEEE International Conference.
3. BERITELLI. F, 1998 A robust voice activity detector for wireless communications using soft computing / F. Beritelli, S. Casale, A. Cavallaro // IEEE Journal on Selected Areas in Communications. V. 16, № 9. – P. 1818–1829.
4. CHO. Y, D, 2001 Analysis and improvement of a statistical model-based voice activity detector / Y. D. Cho, A. Kondoz // IEEE Signal Processing Letters. – V. 8, № 10. – P. 276–278.
5. DORTIZ. P, LUISA. F, VILLA, CARLOS SALAZAR and QUINTERO. O, L, **2016** A simple but efficient voice activity detection algorithm through Hilbert transform and dynamic threshold for speech pathologies. Journal of Physics: Conference Series, Volume 705, conference 1.
6. EVERTELETSKAYA. E, SAKHNOV. K, **2010** Voice Activity Detection for Speech Enhancement Applications. Acta Polytechnica Vol. 50 No. 4.
7. GAZOR. S, 2003 A soft voice activity detector based on a Laplacian-Gaussian model, IEEE Transactions on Speech and Audio Processing. – V. 11, № 5. – P. 498–505.
8. GOLDMAN-EISTER. F, 1972 Pauses, clauses, centences / F. Goldman-Eister // Language and Speech. – V. 15, № 3. – P. 103–113.
9. IVAN. J, TASHEV., SEYEDMAHDAD MIRSAMADI., **2016** DNN-based Causal Voice Activity Detector. Published In Information Theory and Applications Workshop, University of California - San Diego, February 2.
10. JING PANG. **2017** Spectrum energy based voice activity detection. Computing and Communication Workshop and Conference (CCWC), IEEE 7th Annual.
11. LUIS FERRER-ARNAU, JUIAN-GONAZALAZ, VICENC PARISI-BARADAD, **2013** Operators to calculate the derivative of digital signals. 19th IMEKO TC 4 Symposium and 17th IWADC Workshop Advances in Instrumentation and Sensors Interoperability, July 18-19, Barcelona, Spain.
12. MAGDALENA IGRAS-CYBULSKA, MARCIN WITKOWSKI, **2016** Structure of pauses in speech in the context of speaker verification and classification of speech type.

EURASIP Journal on Audio, Speech, and Music Processing.

13. MCKINELY. B, L, 1997 Model based speech pause detection / B. L. McKinley, G. H. Whipple // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’97). – Munich, Germany. – V. 2. – P. 1179–1182.
14. MD SAHIDULLAH. **2012** Comparison of Speech Activity Detection Techniques for Speaker Recognition. IEEE 2012.
15. OMID GHAHABI, WEI ZHOU, VOLKER FISCHER, **2018** A robust voice activity detection for real-time automatic speech recognition. EML European Media Laboratory GmbH**.**
16. PWINT. M, **2006** Speech / Nonspeech Detection Using Minimal Walsh Basis Functions / M. Pwint, F. Sattar // EURASIP Journal on Audio, Speech, and Music Processing. – V. 2007. – P. 3–12.
17. REZAYEE. A, 2001 An adaptive KLT approach for speech enhancement / A. Rezayee, S. Gazor // IEEE Transactions on Speech and Audio Processing. – V. 9, № 2. – P. 87–95.
18. SHEIKHZADEH. H, 1995 Real-time implementation of HMM-based MMSE algorithm for speech enhancement in hearing aid applications / H. Sheikhzadeh, R. L. Brennan, H. Sameti // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’95). – Detroit, Mich, USA,– V. 1. –P. 808–811.
19. SOHN. J, 1999 A statistical model-based voice activity detection / J. Sohn, N. S. Kim, W. Song // IEEE Signal Processing Letters. – V. 6, № 1. – P. 1–3.
20. VANOOP. V, RAO. P, V, **2016** Speech Enhancement Using Hybrid Optimization Technique. Procedia Technology, Volume 25, Pages 5-11.
21. WEIQING ONG**,** ALANWEE TAN**, 2017** Real-time Voice Activity Detector Using Gammatone Filter and Modified Long-Term Signal Variability. ICISPC 2017 (Proceedings of the International Conference on Imaging, Signal Processing and Communication), Penang, Malaysia — July 26 - 28, Pages 113-117.
22. YANNA MA, AKINORI NISHIHARA, **2013** Efficient voice activity detection algorithm using long-term spectral flatness measure. EURASIP Journal on Audio, Speech, and Music Processing.
23. Альтман, Я. А. Тенденции развития физиологии слуха / Я. А. Альтман, С. Ф. Вайтулевич, И. А. Вартанян, И. Г. Андреева, Е. С. Малинина // Успехи физиологических наук. – **2005**. – Т. 36, № 1. – С. 3–23.
24. Варшавский, Л. А. Исследование формантного состава и некоторых других характеристик звуков русской речи / Л. А. Варшавский, И. М. Литвак // Проблемы физиологической акустики. – 1955. – Т. 3. – С. 5–17.
25. Николлс, Д. От нейрона к мозгу / Д. Николлс, Р. Мартин, , Б. Валлас, П. Фукс ; пер. с англ. П.

М. Балабана, А. В. Галкина, Р. А. Гиниатуллина, Р. Н. Хазипова, Л. С. Хируга. – М. : Едиториал УРСС, 2003. – 672 с.

26. Утробин, В. А. Информационные модели системы зрительного восприятия для задач компьютерной обработки изображений : учебное пособие / В. А. Утробин. – Н. Новгород : Изд-во НГТУ, 2001. – 234 с.
27. Утробин, В. А. Компьютерная обработка изображений. Принятие решений в пространстве эталонов : учебное пособие / В. А. Утробин. – Н. Новгород : Изд-во НГТУ, **2004.** – 221 с.
28. Шейкин, Р. Л. К анализу механизмов возникновения пауз в речи / Р. Л. Шейкин // Механизмы речеобразования и восприятия сложных звуков. – М.-Л. : Наука, 1966. – С. 31–44.
29. Элементы теории биологических анализаторов / под ред. Н. В. Позина. – М. : Наука. Главная редакция физико-математической литературы, 1978. – 360 с.