

## خوارزمية تحديد المقاطع النشطة والصامتة في الإشارة الكلامية

د. م حسان محمد أحمد<sup>(1)</sup>

### المخلص

هدف هذا البحث إلى دراسة خوارزمية تحديد المقاطع النشطة والصامتة وتطويرها (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 إمكانية استخدامها لحل المسألة المحددة والمفروضة في مجال معالجة الإشارة الكلامية (نظم تعرّف على المتحدث باستخدام بصمة صوته).

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

<sup>(1)</sup> مدرس في قسم هندسة الاتصالات والشبكات - كلية هندسة الحاسوب والمعلوماتية والاتصالات - الجامعة السورية الخاصة.

## Algorithm of isolating the pauses in a speech signal

Dr. Hassan M. Ahmad<sup>(1)</sup>

### Abstract

The article considers an algorithm of isolating the active segments and 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 capability of the offered algorithm a for solving the specific problem in speech processing area person identification through (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)

## المقدمة

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

في هذا العمل، نعرّف المقطع الصامت (pause) بأنّه المكان في الإشارة الكلامية الذي لا يوجد فيه نشاط كلامي، أي خال من الكلام.

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

عادةً يتحدد الطول الزمني للمقطع الصامت من خلال السبب المؤدي الى تشكيله. تتشكل المقاطع الصامتة في 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]. في أثناء عملية القراءة، يكون طول المقطع اللفظي الصامت (linguistic pause) عند حدود سلسلة من الكلمات المركبة (syntagmatic) أو عند نهاية مقطع صرفي (morpheme) لا يتجاوز 0.75s، في حين يتغير بين الجُمَل من 0.5 إلى 1.5s [8-12-28].

لذلك استخدمنا في هذا العمل احتمالية أنّ طول المقطع الصامت يمكن أن يتغير بين 0.1 إلى 1.5s.

تعدّ مسألة وجود الضجيج (noise) في الإشارة الكلامية الأولية إحدى أهم المشكلات التي تظهر في أثناء عملية التقطيع. بناءً على ذلك فمن 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 الرّكائبيّة (stapedius muscle) للأذن الوسطى يسمح بزيادة صلابة سلسلة العظام السمعية، ممّا يؤدي إلى انخفاض الطاقة الناجمة عنهم، ويسمح بتعويض تشويشات الكثافة العالية.

## هدف البحث

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

### 1. استعراض الخوارزميات المتوافرة

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

يقدم الباحث Omid Ghahabi وزملاؤه في عملهم [15] تقنية تحديد النشاط الصوتي المنيع لأجل نظم التعرف التلقائي الى الكلام في الزمن الحقيقي المبنية على أساس إحصائيات وخوارزمية Baum-Welch .

- قدّم الباحث Gazor في عمله [7] خوارزمية تحديد نشاط الكلام المبنية على استخدام تحويل جيب التمام المتقطع (discrete cosine transform)، إذ صمّمت الخوارزمية مع الأخذ بالحسبان حقيقة أنّ قانون توزيع الكلام والضجيج يتوافق مع نماذج غاوص ولايبلاس (Gaussian and Laplace models).
- قدّم الباحث Sheikhzadeh في عمله [18] خوارزمية تقطيع الإشارة على أساس حساب دالة الارتباط الذاتي للإشارة. طوّرت هذه الخوارزمية لنظم الزمن الحقيقي وقد نفّذت على منصة معالجة الإشارة الرقمية (DSP-platform) لأجل تحسين الكلام في الأجهزة السمعية.
- قدّم الباحث Rezayee في عمله [17] خوارزمية تحديد المقاطع الصامتة المبنية على أساس تحويل كارنن لوف التكيفي (Karhunen-Loève adaptive transformation).
- قدّم الباحثان Evertletskaيا و Vanoop في أعمالهما [6-20] طرائق وتقنيات تحديد النشاط الصوتي لأجل تحسين الكلام مستخدمين طريقة تحديد النشاط الصوتي المنيع للضجيج، وتقنية التحسين الأمثل الهجين (Hybrid Optimization).
- نرى من الضروري الإشارة إلى أنّ الطرائق والخوارزميات المتوفرة لتحديد المقاطع الصامتة في الإشارة الكلامية تسعى دوماً في تحسين واحد من البارامترات الآتية: دقة تحديد المقطع الصامت، والتأخر 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]:
- قدّم 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، إذ استخدم 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 الإشارة الكلامية مستخدماً الطريقة التكيفية للتحليل المتكامل لوظائف الحالة الداخلية للمتكلم.
- قدّم الباحث Ortiz في عمله [5] خوارزمية بسيطة وفعالة لتحديد النشاط الصوتي مستخدماً تحويل هيلبرت (Hilbert transform) والعتبة الديناميكية لأعراض الكلام ومشكلاته.
- قدّم الباحث Beritelli في عمله [3] طريقة لتقطيع الإشارات مبنية على أساس نظرية تعرّف الكلام ونظرية المنطق المبهيم (fuzzy logic). يمكن استخدام هذه الطريقة لمعالجة الإشارات المضججة.
- قدّم الباحث McKinley في عمله [13] خوارزمية تحديد النشاط الصوتي على أساس الحدّ الأعظمي للمعلومات الاستدلالية (posteriori information)، ومعيار (criterion) نيمان-بيرسون (Neyman-Pearson).
- قدّم الباحثان Sohn و Cho في أعمالهما [4-19] خوارزمية تحديد المقاطع النشطة والصامتة المبنية على فرضية أنّ تحويل فورييه المتقطع (discrete Fourier transform) للكلام والضجيج يعدّ متغيرات عشوائية غوسية (Gaussian) مستقلة ومتقاربة.

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 أن هذا المقطع هو مقطع صامت، ومن ثم حذفه (عذله) من عملية المعالجة اللاحقة.

(4) إذا كانت اشتقاقات المقطع ذي التسلسل  $i$  غير مساوية للصفر، فإننا نستمر في تقسيم المقطع  $i$  الذي يجري تحليله إلى مقاطع جزئية أخرى، ومن ثم نطبق الخطوات السابقة 2 و 3 على كل من المقاطع الجزئية الناتجة.

تتهي الخوارزمية عملها عندما يصنّف 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 تصنّف، والتي اشتقاقاتها لا تساوي الصفر، فإنه يتم تسجيل مثل هذه المقاطع على أنها مقاطع كلامية نشطة (غير صامتة).

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

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

اذ:  $F$  - تردد أخذ العينات.

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

$$= 10 \log_{10} \frac{P_s}{P_v}$$

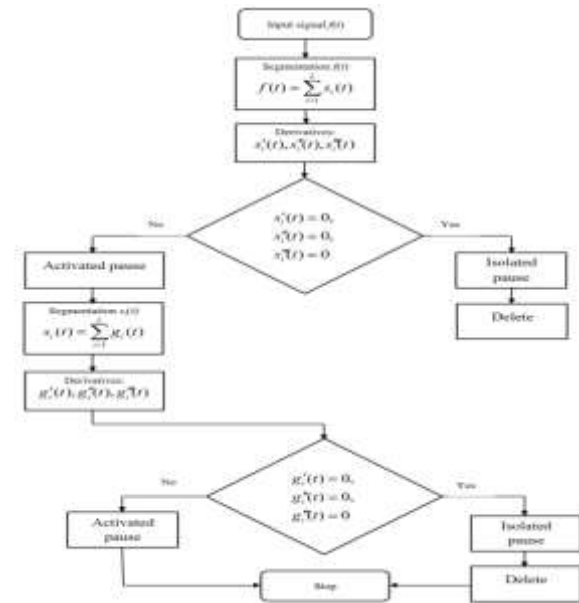
اذ:

$s(n)$  - 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.

## 2. الخوارزميات المقترحة

4.1 خوارزمية تحديد المقاطع الصامتة في الإشارة غير المضججة.

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

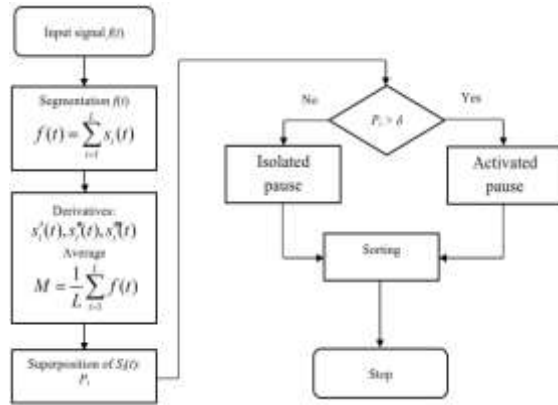


الشكل (1) المخطط الانسيابي لخوارزمية تحديد المقاطع الصامتة في الإشارة غير المضججة  
 توصيف المخطط الانسيابي المبين في الشكل (1):

## 4.2 خوارزمية تحديد المقاطع الصامتة في الإشارة المضججة

تسمح الخوارزمية التي وصفت سابقاً بتقطيع الإشارة غير المضججة. لكن في الظروف الحقيقية لا يمكن عملياً أن نجد مثل هذه الإشارات بسبب وجود التشوهات في خطوط الاتصالات والأجهزة المستخدمة، ووجود الضجيج من البيئة المحيطة. ومن ثم فإن خوارزمية تحديد المقاطع الصامتة وعزلها لا بد أن تأخذ بالحسبان وجود الضجيج في الإشارة.

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



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

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 تراكم (superposition) قيم البرامترات كلها الناتجة لكل مقطع، فإذا كانت نتيجة حساب التراكم للمقطع ذي التسلسل  $i$  أكبر من قيمة عتبة معينة،

فيمكن الافتراض بأن 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]).  
تتهي الخوارزمية عملها عندما يصنف كل مقطع من مقاطع الإشارة.

## 5. التجربة الحاسوبية

أجريت التجارب الحاسوبية وسجلت 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، وعمق الترميز (coding depth) 16bit، بطول 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)) يختار الطول الناتج بمنزلة عدد، يكون أساً للعدد 2. ومن ثم نختار الطول الأصغري للمقطع  $L=128$  عينة (8ms).  
تجدر الإشارة إلى أن النظام السمعي يُقيّم التغيرات التي تحصل خلال مدة زمنية أقل من 1ms في النبضة الصوتية [29].

في أثناء اختبار الخوارزمية المقترحة على الإشارة الكلامية المضججة استخدمت الأنواع الآتية من الضجيج الجمعي:

### (1) النظرية (theoretical):

- ضجيج جمعي طبيعي (غوسي) (Additive normal noise (Gaussian))؛
- ضجيج جمعي منتظم (Additive uniform noise) (noise).

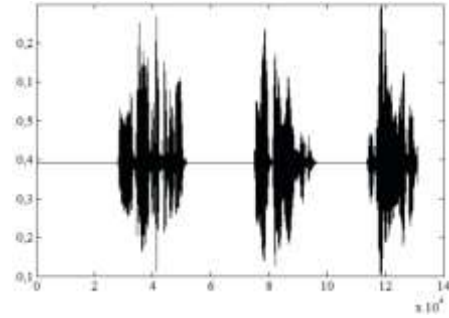
### (2) الحقيقية (real):

- ضجيج طريق سريع (اتوستراد).
- ضجيج الشارع.
- ضجيج محرك السيارة.

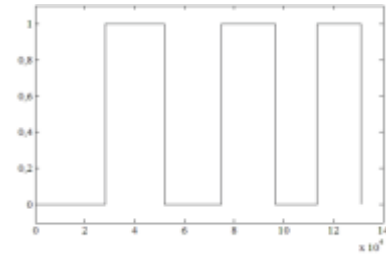
يعدّ ضجيج الطريق السريع (اتوستراد) شبيهاً بالضجيج الطبيعي. يتجلى الفرق بينهما في حقيقة أنّ ضجيج الطريق السريع يحتوي على مقاطع تزداد فيها الكثافة (intensity) عند اقتراب السيارة من أدوات تسجيل الصوت.

بيّن الجدول (1) نتائج التجارب الحسابية التي أجريت على الإشارة الكلامية المضججة بأنواع مختلفة من الضجيج. يشير العمود err1 إلى الخطأ من النوع الأول الذي يبيّن النسبة المئوية لكمية المقاطع الصامتة التي حدّدت بشكل كاذب (false pause)، في حين يشير العمود err2 إلى الخطأ من النوع الثاني الذي يبيّن النسبة المئوية لكمية المقاطع الصامتة التي تم تجاوزها (skipped pauses).

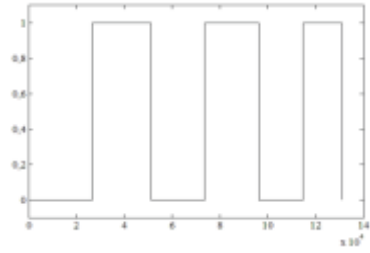
بيّن الجدول (2) نتائج تقييم جذر مربع متوسط انحراف الضجيج (Root-Mean-Square Deviation, RMSD).



الشكل (3) الإشارة الكلامية الأولية

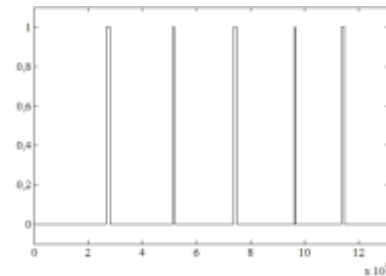


الشكل (4) التقطيع اليدوي للإشارة الكلامية

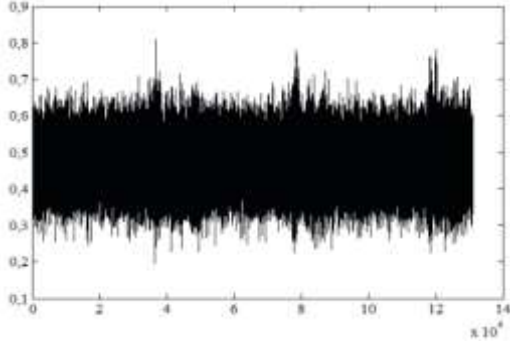


الشكل (5) التقطيع الآلي للإشارة الكلامية

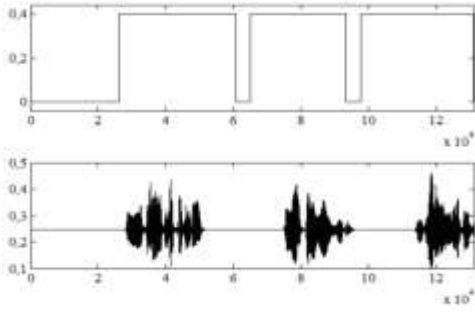
بيّنت التجارب الاختبارية التي أجريت على الإشارة الكلامية الأولية باستخدام الخوارزمية المقترحة (الشكل (6)) أن الخطأ من النوع الأول (خطأ تحديد المقطع الصامت) قد بلغ 0.02، في حين بلغ الخطأ من النوع الثاني (تجاوز المقطع الصامت) 0.017، وقد بلغ زمن معالجة الإشارة الأولية 0.18s.



الشكل (6) الفرق بين التقطيع اليدوي والآلي



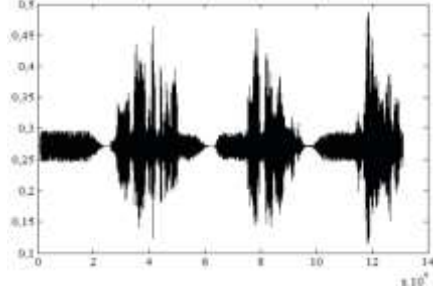
(a)



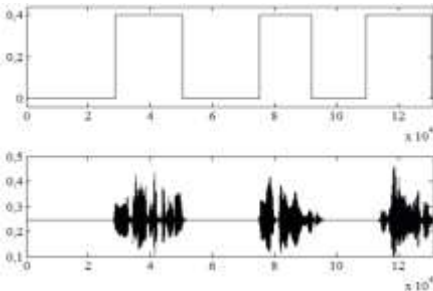
الشكل (8) إشارة مضججة بضجيج طبيعي

(SNR = 0dB)

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



(a)



(b)

الشكل (9) إشارة مضججة بضجيج محرك سيارة

(SNR = 20dB)

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

الجدول (1) نتائج عمل الخوارزمية

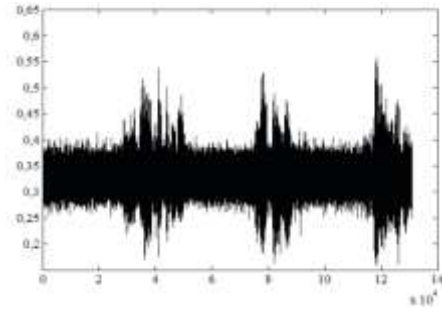
نوع الضجيج										SNR
ضجيج محرك سيارة		ضجيج الشارع		ضجيج التوتربك		ضجيج منظم		ضجيج طبيعي		
err2	err1	err2	err1	err2	err1	err2	err1	err2	err1	
5	0	5.8	0	5.6	5	6	0	6	0	30 dB
6	0	4.2	0	5.9	4.3	6.7	0	6	0	20 dB
3.4	1.5	3.3	5.3	3.9	3.7	3.8	10	7.1	0	10 dB
3.5	14.8	1.5	20.1	4.4	3.8	1.9	32.4	7.8	1.5	5 dB
3.7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

الجدول (2) خصائص الضجيج (RMSD للضجيج)

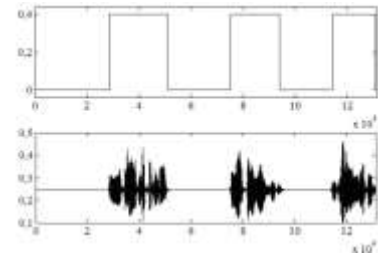
نوع الضجيج					SNR
ضجيج محرك سيارة	ضجيج الشارع	ضجيج التوتربك	ضجيج منظم	ضجيج طبيعي	
0.00089	0.00083	0.00091	0.00405	0.00171	30 dB
0.00445	0.00417	0.00454	0.01327	0.00559	20 dB
0.01423	0.00835	0.01362	0.04034	0.02021	10 dB
0.02313	0.01669	0.02088	0.05762	0.03495	5 dB
0.03503	0.03004	0.84024	0.11537	0.05013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

تبيّن الأشكال (7) و (8) و (9) إشارة كلامية أولية

ونتيجة تقطيعها.



(a)



(b)

الشكل (7) إشارة مضججة بضجيج طبيعي

(SNR = 10dB)

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



## 5. الاستنتاجات

يمكننا أيضاً تمييز الخوارزمية المقترحة مقارنة

بالخوارزميات الأخرى المتوفرة بالنقاط الإيجابية الآتية:

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

تتجلى نتائج دراسة الخوارزمية المقترحة المناعية للضجيج في الآتي:

(1) تنخفض دقة التقطيع عند زيادة الطول الزمني للإشارة التي تتم معالجتها. أحد الحلول الممكنة لهذه المشكلة هو المعالجة المحددة لمقاطع الإشارة بطول 2-3 ثانية.

(2) إذا قارنا دقة تحديد حدود المقاطع  $S_1$  و  $S_2$  (اذ:  $A_1$  المطال الأعظمي للمقطع  $S_1$  و  $A_2$  المطال الأعظمي للمقطع  $S_2$ ؛  $A_1 \square A_2$ ) فإن دقة تحديد حدود المقطع  $S_1$  تكون أقل من دقة تحديد حدود المقطع  $S_2$ ، وذلك بسبب اختلاف مستوى تأثير الضجيج المرفق للإشارة في مقاطعها المختلفة المطالات.

(3) إذا قارنا دقة تحديد حدود المقاطع  $S_1$  و  $S_2$  (اذ:  $L_1$  طول المقطع  $S_1$  و  $L_2$  طول المقطع  $S_2$ ؛  $L_1 \square L_2$ ) 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

في هذا العمل درست وطوّرت خوارزمية تقطيع الإشارة الكلامية إلى مقاطع نشطة الكلام وصامتة. قدّم هذا العمل خوارزمتين: الأولى مخصصة لتقطيع الإشارات المضججة، والثانية لتقطيع الإشارات غير المضججة. عند تصميم الخوارزمتين المذكورتين تم الأخذ بالحسبان طرائق مبنية على أساس الآليات المماثلة التي تعمل في نظام الإدراك البصري.

## REFERENCES

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 GHAAHABI, 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.
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, sentences / 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.

- 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.
- 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. Фуке ; 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 :

Received	2018/07/09	إيداع البحث
Accepted for Publ.	2018/11/26	قبول البحث للنشر