مجلة جامعة تشرين للدراسات و البحوث العلمية _ سلسلة العلوم الهندسية المجلد (24) العدد (12)2002
Tishreen University Journal for Studies and Scientific Research- Engineering Sciences Series Vol (24) No (12) 2002

تحسين إشارة الصوت باستخدام مرشح كالمان الممتد باعتماد الشبكات العصبية

الدكتور زهير وقاف *

(قبل للنشر في 2002/4/22)

□ الملخّص □

إن مسألة تخفيض الضجيج في إشارات الكلام تعتبر ذات أهمية كبيرة في التطبيقات المختلفة بدءاً من نظم التعرف على الصوت، إلى تحسين الاتصالات عن بعد في مجال الطيران المدني والعسكري وإجراء المؤتمرات عن بعد، والاتصالات الخليوية.

إن الهدف هو تحسين نوعية إشارة الكلام المستقبلة أو زيادة انجابيتها. وعادةً ما تستخدم الطرق الطيفية لتحقيق هذا الغرض في مثل هذه التطبيقات. ولكنها غالباً ما تترافق بتشوهات في المجال السمعي.

سوف نعرض في هذا المقال طريقة للترشيح في مجال الزمن غير خطية تعرف باسم مرشح كالمان ثنائي الامتداد ونبين أنها تمتلك مزايا هامةفي مجال ترشيح إشارة الصوت من الضجيج المتغير والضجيج الملون، إن المنهج المتبع يستخدم نموذج متنبئ مترافق مع مرشح كالمان الممتد وذلك باعتماد الشبكات العصبية الصنعية في المجال العقدي من خلال عرض خوارزمية المرشح مع بعض النتائج العملية.

* مدرس في قسم الهندسة الالكترونية - كلية الهندسة الميكانيكية والكهربائية – جامعة تشرين اللاذقية – سورية.

مجلة جامعة تشرين للدراسات و البحوث العلمية _ سلسلة العلوم الهندسية المجلد (24) العدد (12)2002
Tishreen University Journal for Studies and Scientific Research- Engineering Sciences Series Vol (24) No (12) 2002

Speech Signal Enhancement Using Neural Dual Extended Kalman Filtering

Dr. zouhir wakkaf *

(Accepted 22/4/2002)

□ ABSTRACT □

The removal of noise from speech signals has important applications ranging from speech recognition systems, to enhancement of tele-communications in aviation, military, teleconferencing, and cellular environments. the goal is either to improve the perceived quality of the speech, or to increase its intelligibility. Spectral techniques are commonly used in these applications, but frequently result in audible distortion of the signal. A nonlinear time-domain method called dual extended Kalman filtering(DEKF) is presented, that demonstrates significant advantages for removing nonstationary and Colored noise from speech. the approach uses a predictive model in conjunction with an extended Kalman filtring based on neural networks in complex domain. This paper describes the algorithm and some experimental results.

***Department of Electronic Engineering- Faculty of Mechanical and Electrical Engineering – Tishreen University – Lattakia – Syria.**

مقدمة:

توجد طرق مختلفة لتحسين إشارة الصوت تستخدم تقنيات متنوعة منها(الفصل الطيفي، الربط الجزئي بفضاء الإشارة، بالإضافة إلى طرق أخرى في مجال الزمن) **[1]** . إن هذه الطرق غالباً ما تترافق مع تشوه مسموع في الإشارة. ولا تعتبر جيدة كفاية ضمن وجود الضجيج الاعتيادي في الاوساط الحقيقية.

إن الطرق الحديثة للترشيح بوساطة الشبكات العصبية التي تعمل في المجال الحقيقي تستخدم لتدريب الشبكة مجموعة معطيات مأخوذة من إشارة كلام نظيفة (غير مشوشة) تمثل الإشارة الهدف. إن هذه الطرق تعتبر فعالة بالنسبة للاشارات التي استخدمت في تدريب الشبكة. لكن خاصية التعميم لها(مقدرة الشبكة على ترشيح إشارات لم تدرب عليها سابقاً) تعتبر ضعيفة بالنسبة للاشارات المأخوذة من أوساط حقيقية بعيدة عن مجموعات التدريب، حيث يلاحظ تغير في مستوى الاشارة وكذلك مستوى الضجيج. بالإضافة لذلك فإن نماذج الشبكة العصبية في هذه الطرق لاتأخذ بالاعتبار وبشكل كامل الطبيعة المتغيرة للإشارات الصوتية.

في المدخل المقترح تم افتراض وجود الاشارة المشوشة فقط(بدون معرفة الإشارة النظيفة). كما تم تدريب شبكة عصبية في المجال العقدي على إشارة الكلام المشوشة للحصول على نموذج ديناميكي لترشيح إشارة الكلام من الضجيج.

النموذج غير الخطي لإشارة الكلام:

يمكن نمذجة اشارة الكلام المشوشة $y(k)$ كتغير مترابط غير خطي لاشارة الكلام مع ضجيج جمعي:

$$x(k) = F(x(k-1), \ldots\ldots, x(k-M), W) + v(k) \qquad (1)$$

$$y(k) = x(k) + n(k) \qquad (2)$$

حيث $x(k)$ تمثل إشارة الكلام الحقيقية والمصحوبة بضجيج العملية $v(k)$.

$f(.)$ تابع غير خطي لقيم $x(k)$ ومرتبطة بالبارامتر W .

حيث نفترض أن إشارة الكلام ثابتة فقط ضمن فترات زمنية صغيرة. إن الاشارة $y(k)$ قابلة للمراقبة وتحتوي على ضجيج جمعي $n(k)$. إن القيمة المقّدرة لاشارة الكلام $\hat{x}(k)$ يمكن إيجادها بوساطة عملية تقدير وذلك اعتماداً على القيم المراقبة حتى اللحظة k . أو بوساطة عملية تنعيم اعتماداً على كل القيم المراقبة السابقة والمستقبلية.

إن المقّدر المثالي الذي يتلقى القيم المراقبة $Y(k) = \{y(k), y(k-1), \ldots. y(0)\}$ يعطى بالعلاقة:
$E[x(k) \mid Y(k)]$.

إن الطريقة المباشرة لتقدير هذه القيمة هي بتدريب شبكة عصبية على مجموعة معطيات نظيفة بحيث تكون فيها الاشارة $x(k)$ الحقيقية تمثل هدف الشبكة (إشارة الخرج). ولكن المشكلة هي أننا لانعرف الاشارة $x(k)$ النظيفة وإنما المشوبة بالضجيج فقط غير أنها قابلة للقياس والمراقبة. فإذا استخدمنا $y(k)$ بدلاً من $x(k)$ كإشارة هدف لتدريب الشبكة العصبية فإن المقّدر سوف يعطي الاشارة نفسها ولن نحصل على تقدير للإشارة الحقيقية أي أن $E[x(k) \mid Y(k)]$. ولحل هذه المسألة سوف نفرض أن التابع $f(.,.)$ يكافئ نماذج شبكات عصبية ذات انتشار أمامي. وسوف نقوم بحساب التقدير الثنائي لكل من الحالات $\hat{x}$ والأوزان $\hat{W}$ اعتماداً على منهج مرشح كالمان الممتد.

نموذج فضاء الحالة باعتماد الشبكات العصبية:

عند صياغة مسألة التقدير الثنائي مع التمثيل بطريقة فضاء الحالة يمكن استخدام طرق ترشيح كالمان وذلك لانجاز عملية التقدير بشكل فعال وباسلوب عودي(Recursive). حيث أنه في كل لحظة زمنية فإن مرشح كالمان يعطي التقدير الأمثلي وذلك عن طريق ربط القيمة السابقة مع القيم المراقبة الجديدة. اقترح Connore استخدام مرشح كالمان الموسع مع شبكة عصبية وذلك لاجراء تقدير الحالة فقط [2].

بعد ذلك فإن Puskorlous & Feldkamp وغيرهم اقترحوا تقدير الوزن بطريقة فضاء الحالة وذلك لزيادة فعالية تدريب الشبكة العصبية بوساطة مرشح كالمان [3].

سوف نهتم الان بتوسيع هذه الافكار لتشمل تقدير (Estimation)كالمان الثنائي لكل من الحالات $\hat{x}$ والأوزان $\hat{W}$ وذلك لزيادة فعالية التنبؤ والتقدير والتنعيم وسنستخدم اشارة الصوت كمثال هنا.

لتطبيق مرشح كالمان الموسع (EKF) سنكتب المعادلات (1) و (2) بصيغة فضاء الحالة:

$$X(k) = F[X(k-1)] + B\ v(k) \qquad (3)$$

$$y(k) = C\ X(k) + n(k) \qquad (4)$$

حيث:

$$X(k) = \begin{bmatrix} x(k) \\ x(k-1) \\ \vdots \\ x(k-M+1) \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$F[X(k)] = \begin{bmatrix} f(x(k), \ldots\ldots, x(k-M+1), W) \\ x(k) \\ \vdots \\ x(k-M+2) \end{bmatrix}$$

و $C = B^T$. فإذا كان النموذج خطياً عندها فإن التابع $f(X(k))$ يأخذ الشكل $W^T . X(k)$ وعندها يمكن كتابة التابع $F[X(k)]$ من الشكل $A.X(k)$ حيث A مصفوفة بالشكل القانوني القابل للمراقبة.

في البداية سوف نفرض أن حدود الضجيج $v(k)$ و $n(k)$ تمثل ضجيج أبيض ذو تشتت s_v^2 و s_n^2 على التوالي.

تقدير الحالة:

إذا كان النموذج خطياً وذو بارامترات معروفة فإن خوارزمية مرشح كالمان (KF) يمكن تطبيقها مباشرة لتقدير الحالة [4]. فعند كل خطوة زمنية فإن المرشح يقوم بحساب متوسط المربعات الخطي للقيمة المقدرة $\hat{x}(k)$ ولقيمة

التنبؤ $\bar{\hat{x}}(k)$. وكذلك قيم تباعد الخطأ لكليهما $P_{\hat{X}}$ و $P_{\bar{\hat{X}}}$. ففي الحالة الخطية وإذا كان التوزع الاحصائي غاوصي فإن التقديرات تكون هي القيمة الدنيا للتربيعات الصغرى للتقديرات. وفي حال عدم وجود معلومات أولية عن الاشارة x فسوف تؤول إلى تقديرات بجوار القيمة العظمى.

وفي حال كان النموذج غير خطي فإن KF لايمكن تطبيقه مباشرة ، بل يتوجب تقريب النموذج غير الخطي إلى نموذج خطي عند كل خطوة زمنية. والخوارزمية الناتجة تسمى مرشح كالمان الممتد (EKF) . حيث يقوم وبشكل فعال بتقريب التابع غير الخطي إلى تابع خطي متغير مع الزمن. إن خوارزمية EKF هي كما يلي:

$$\bar{\hat{X}}(k) = F[\hat{X}(k-1), \hat{W}(k-1)] \quad (3)$$

$$P_{\bar{\hat{X}}}(k) = A \cdot P_{\hat{X}}(k-1) \cdot A^T + B \cdot \sigma_v^2 \cdot B^T \quad (4)$$

$$A = \frac{\partial F[\hat{X}, \hat{W}]}{\partial \hat{X}}\Big|_{\hat{X}(k-1)} \quad (5)$$

حيث

$$K(k) = P_{\bar{\hat{X}}}(k) \cdot C^T \; (C \cdot P_{\bar{\hat{X}}}(k) \cdot C^T + \sigma_n^2)^{-1} \quad (6)$$

$$P_{\hat{X}}(k) = (I - K(k) \cdot C) \cdot P_{\bar{\hat{X}}}(k) \quad (7)$$

$$\hat{X}(k) = \bar{\hat{X}}(k) + K(k) \cdot (y(k) - C \cdot \bar{\hat{X}}(k)) \quad (8)$$

حيث أن المشتقات في العلاقة (5) توافق عملية تقريب الشبكة العصبية إلى نموذج خطي عند نقطة العمل. وإذا كانت الأوزان W غير معروفة عندها يمكن استبدالها بالقيم المقدرة $\hat{W}$. أما $K(k)$ فتمثل ربح مرشح كالمان.

تقدير الأوزان:

بما أن نموذج إشارة الصوت غير معروف فإن خوارزمية EKF لايمكن تطبيقها مباشرة، لذلك سنحل هذه المشكلة بإنشاء صيغة مستقلة لتمثيل الأوزان بنموذج فضاء الحالة كما يلي:

$$W(k) = W(k-1)$$

$$y(k) = f(X(k-1), W(k)) + v(k) + n(k) \quad (9)$$

حيث أن مصفوفة الانتقال هي مصفوفة الوحدة. أما الشبكة العصبية $f(X(k-1), W(k))$ فتلعب دور مراقب غير خطي متغير مع الزمن لـ W. معادلات فضاء الحالة لشعاع الأوزان المبينة أعلاه تسمح لنا بتقدير قيمة الأوزان بوساطة EKF الثاني:

$$\bar{\hat{W}}(k) = \hat{W}(k-1) \quad (10)$$

$$P_{\bar{\hat{W}}}(k) = P_{\hat{W}}(k-1) \quad (11)$$

$$K_{\hat{W}}(k) = P_{\bar{\hat{W}}}(k) \cdot H^T(k) \cdot [H(k) \cdot P_{\bar{\hat{W}}}(k) \cdot H^T(k) + \sigma_n^2 + \sigma_v^2]^{-1} \quad (12)$$

$$P_{\hat{W}}(k) = (I - K_{\hat{W}}(k) \cdot H(k)) \cdot P_{\bar{\hat{W}}}(k) \quad (13)$$

$$H(k) = \frac{C \cdot \partial F[\hat{X}, \hat{W}]}{\partial \hat{W}}\Big|_{\hat{W}(k-1)} \quad (14)$$

$$\hat{W}(k) = \bar{\hat{W}}(k) + K_{\hat{W}}(k) \cdot (y(k) - C \cdot F(\hat{X}(k-1), \bar{\hat{W}}(k)) \quad (15)$$

إن استخدام EKF في تقدير شعاع الأوزان يستند إلى مبدأ التربيعات الصغرى العودي، والذي يعتبر طريقة أمثلية في الزمن الحقيقي ومن الدرجة الثانية.نذّكر بأنه عندما تكون x غير معروفة فيجب استبدالها في مرشح الأوزان بالقيمة التقديرية $\hat{x}$.إن عملية تقريب نموذج الشبكة العصبية إلى نموذج خطي بوساطة المعادلة (14) تؤخذ كاشتقاق جزئي أو كلي وفي كلا الحالتين تتطلب زمناً كبيراً في تدريب الشبكة وللسهولة نهمل تبعية $\hat{X}(k-1)$ لـ $\hat{W}$. وبالتالي يصبح تقريب الشبكة كتقريب ستاتيكي حيث بينت النتائج أن الاشتقاق الجزئي لايعطي نتائج أفضل بكثير من حالة الاشتقاق الستاتيكي[5].

التقدير المزدوج لكل من الحالة والأوزان:

إن ضرورة وأهمية وجود خوارزمية مرشح كالمان ثنائي الامتداد (DEKF) هو تشغيل مرشح شعاع الحالة ومرشح شعاع الأوزان على التوازي كما يبين الشكل (1).

حيث يتم تعديل لحظي لكل من الحالة x(k) والوزن W. ففي كل خطوة زمنية فإن التقدير الحالي لـ x يستخدم من قبل مرشح الوزن كما أن التقدير الحالي للوزن W يستخدم من قبل مرشح الحالة.

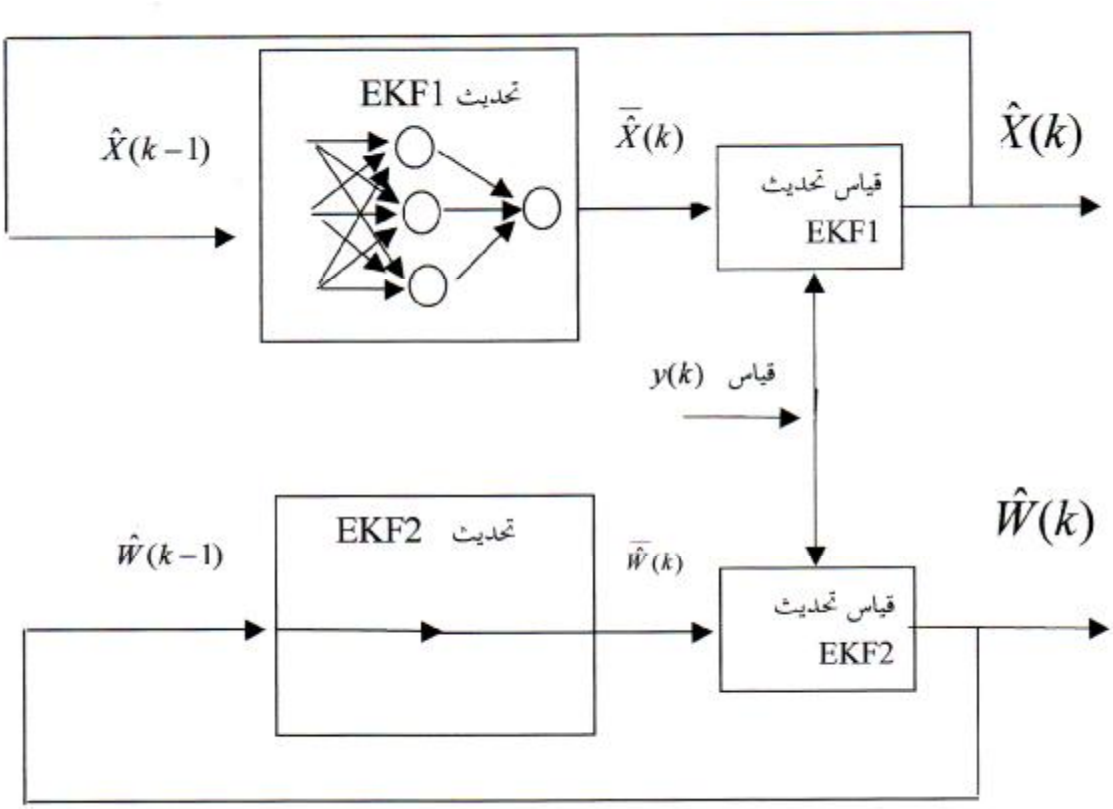
فعندما تكون مجموعة معطيات الدخل محدودة فإن الخوارزمية سوف تعالج المعطيات حتى تتم عملية تقارب الأوزان[6].

معالجة إشارة الصوت:

لمعالجة إشارة كلام مشوبة بالضجيج(حيث افترضنا أن الضجيج أبيض) تم تطبيق الطريقة على تتابع نوافذ من الاشارة بعرض 64 msec (512 نقطة) حيث يفترض أن الاشارة ثابتة خلال النافذة، مع نافذة جديدة تبدأ كل 8 msec.

تم استخدام نافذة هامينغ المعدلة لتوضيح المعطيات في مركز النافذة وتهميشها في الأطراف وذلك عند تجميع النوافذ المتراكبة. أيضاً تم استخدام النوافذ في عملية تقدير الأوزان.

في عملية الترشيح استخدمت شبكة عصبية ذات انتشار أمامي في المجال العقدي لها (10) مداخل و (4) عقد في الطبقة المخفية وعقدة في طبقة الخرج[7]. إن النتائج في الاشكال تم احتسابها بافتراض أن كلاً من s_v^2 و s_n^2 معروفين. في النتيجة تحسنت القيمة المتوسطة لنسبة الاشارة إلى الضجيج SNR بمقدار 9 dB. إن نتائج تطبيق EKF الثنائي لاشارات الكلام مبين على الشكل (2 -a,b,c,d) . نذكر هنا أنه تم بناء نموذج المرشح مع الشبكة العصبية بوساطة برنامج الـ MATLAB وتم اختيار إشارة الصوت كمثال للترشيح بوساطة مرشح كالمان نظراً لأهمية التطبيقات المختلفة التي تقوم على إشارة الكلام ومنها نظم التعرف على الصوت ونظم التعرف على المتحدث وغيرها والتي غالباً ماتترافق مع إشارة ضجيج.



الشكل(1) يبين مرشح كالمان ثنائي الامتداد DEKF حيث أن EKF1 و EKF2 هما مرشحا الحالة والاوزان على التوالي. إن مربع "تحديث EKF1"
يشير إلى المعادلات 3-5 والمعادلات 10-11 لـ EKF2 . قياس التحديث
يشير للمعادلات 6-8 لـ EKF1 و 12- 15 لـ EKF2 .

الضجيج الملون:

في تطبيقات الكلام العملية لانستطيع أن نفترض دائماً أن الضجيج المرافق للاشارة هو الضجيج الأبيض. فمن أجل الضجيج الملون فإن معادلات الحالة (3,4) يجب تعديلها قبل تطبيق تقنية مرشح كالمان. وبشكل خاص فإن عملية قياس الضجيج تعطي معادلات الحالة التالية:

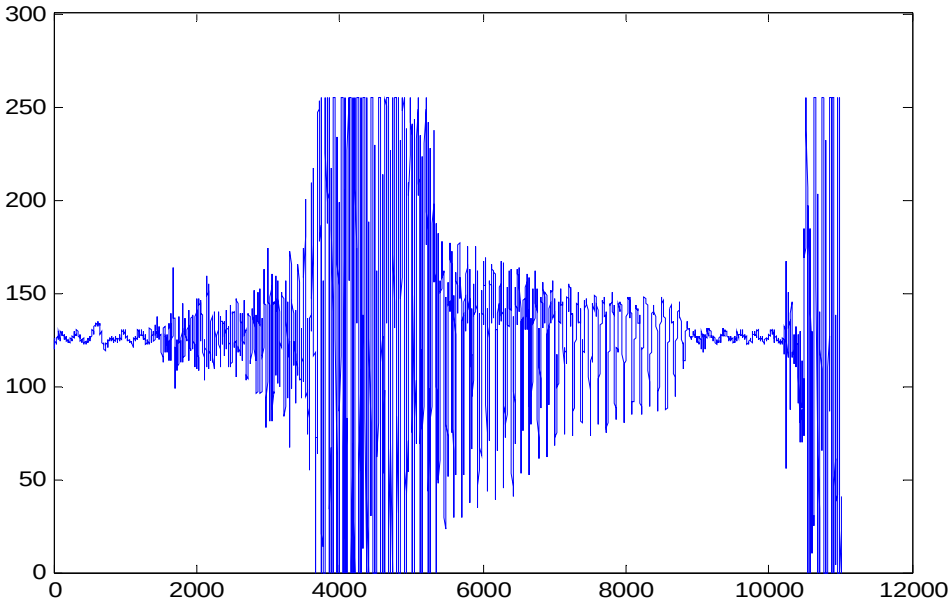
$$n(k) = A_n \cdot n(k-1) + B_n \cdot v_n(k) \qquad (16)$$

$$\bar{n}(k) = C_n \cdot n(k) \qquad (17)$$

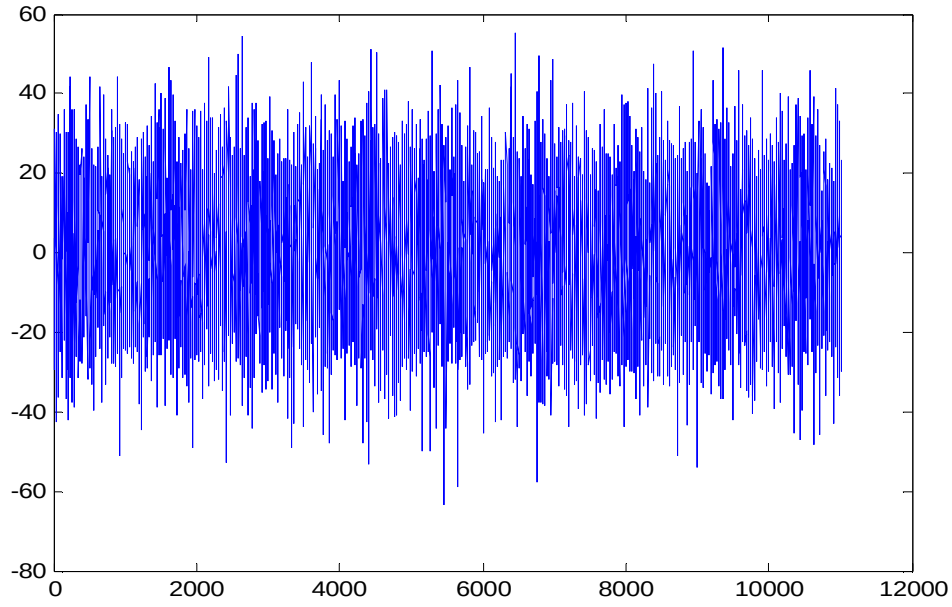
حيث $n(k)$ هو شعاع من القيم المؤخرة لـ $\bar{n}(k)$.

$v_n(k)$ الضجيج الأبيض.

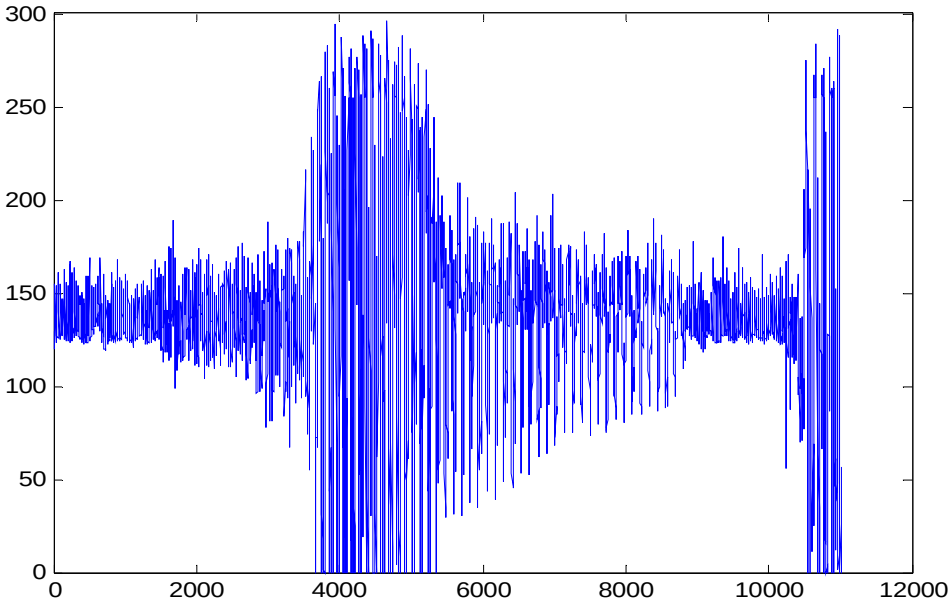
A_n مصفوفة النقل بصيغة قانونية قابلة للتحكم.



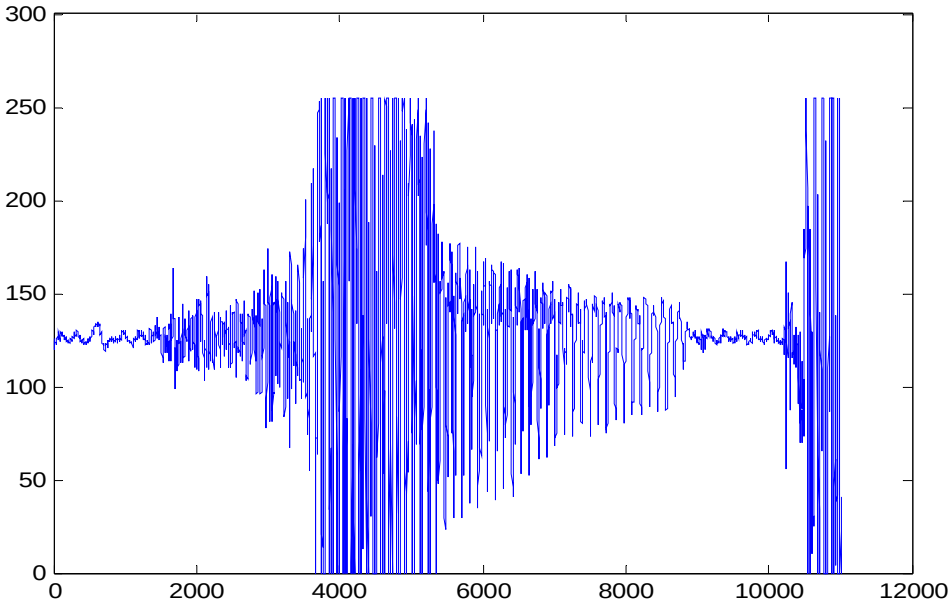
a- الإشارة النظيفة



b- إشارة الضجيج



c- الإشارة المشوبة بالضجيج



d- الإشارة على خرج المرشح

الشكل (2) يبين عملية ترشيح اشارة كلام من الضجيج المرافق باستخدام (DEKF).

B_n , C_n هي مصفوفات كما هو حال B , C في المعادلات (3,4). إن النموذج أعلاه مكافئ لنموذج الضجيج الملون والذي يمكن اقتطاعه من جزء صغير من الاشارة المشوشة حيث لاتوجد ضمنه اشارة كلام حقيقية.

وباعتماد الصيغة أعلاه للضجيج الملون فمن المناسب ضم كلاً من الحالة x(k) والاوزان W(k) مع n(k) وكتابة معادلات الحالة الاجمالية. وبشكل خاص فإن المعادلات (3,4) تصبح من الشكل:

$$\begin{bmatrix} X(k) \\ n(k) \end{bmatrix} = \begin{bmatrix} F[X(k-1)] \\ A_n . n(k-1) \end{bmatrix} + \begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix} \cdot \begin{bmatrix} v(k) \\ v_{n(k)} \end{bmatrix}$$

$$y(k) = [C \quad C_n] \begin{bmatrix} X(k) \\ n(k) \end{bmatrix} \qquad (18)$$

والمعادلات 13 , 12 تستبدل بـ:

$$\begin{bmatrix} W(k) \\ n(k) \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & A_n \end{bmatrix} \cdot \begin{bmatrix} W(k-1) \\ n(k-1) \end{bmatrix} + \begin{bmatrix} 0 \\ B_n \end{bmatrix} \cdot v_n(k)$$

$$y(k) = f(X(k-1), W(k)) + C_n . n(k) + v(k) \qquad (19)$$

وبالتالي يصبح الضجيج المعالج في هذه المعادلات هو الضجيج الأبيض. وبالتالي يمكن استخدام خوارزمية (DEKF)لتقدير الاشارة. نلاحظ من المعادلات أن الضجيج الملون يؤثر ليس فقط على تقدير الحالة ،إنما على تقدير الوزن أيضاً[8].

خاتمة: عرضنا في هذا العمل خوارزمية EKF الثنائي مع نتائج تطبيق هذه الخوارزمية لتحسين اشارة الكلام وذلك بوجود كل من الضجيج غير الثابت والضجيج الملون. ومحاسن هذه الطريقة تظهر جلية عندما يكون الضجيج أبيض وغير ثابت. علماً بأن أدائها في حال وجود الضجيج الملون الثابت أفضل من الطرق المعروفة الأخرى.

المراجع:

.....

1-J.deller. 1993 –Time Processing of speech signals. Macmillan publishing company, U.S.A.

2- J.T.Conner, March 1994 - Recurrent neural networks and robust time series prediction . IEEE Transactions on Neural Networks. U.S.A.

3-G.V.Puskorious and L.A.Feldkamp.1994,5(2). Neural control of nonlinear dynamics systems with Kalman filter trained recurrent networks. IEEE Transactions on Neural Networks,U.S.A.

4 -Carles L. Phillips,H.Troy .1995 – Digital control system analysis and design. 3rd ed. Prentice Hall. U.S.A.

5-P.Werbos. 1992 – Handbook of intelligent control,Fuzzy,Neural,and Adaptive Approaches. U.S.A.

6-G.Goodwin, K.S.Sin.1994- Adaptive filtering prediction and control, Prentice-Hall,inc.U.S.A

7-T.Masters. 1994- Signal and image processing with neural networks. Jon wiley &sons, INC. U.S.A.

8-H.G. Hirsch. 1993 - Estimation of noise spectrum and its application to SNR-estimation and speech enhancement, Technical Report TR-93-012.International Computer Science Institute. U.S.A.