مجلة جامعة تشرين للدراسات و البحوث العلمية _ سلسلة العلوم الهندسية المجلد (28) العدد (2) 2006
Tishreen University Journal for Studies and Scientific Research- Engineering Science Series Vol. (28) No (2) 2006

استخدام الشبكات العصبونية الاصطناعية لإزالة الضجيج من إشارة الكلام

الدكتورة لميس قدسي*
الدكتور بلال شيحا **
حسين الأزكي ***

(قبل للنشر في 2006/10/10)

□ الملخّص □

تحسين نوعية ووضوح إشارة الكلام بإزالة الضجيج منه له مجال واسع من التطبيقات نظراً للحاجة إليه في تحسين أداء أجهزة الاتصالات الصوتية في حالات الضجيج كالملاحة الجوية ومؤتمرات الاتصالات وبيئة الخليوي.

وتعد الشبكات العصبونية الاصطناعية من التطورات المهمة في استخلاص الضجيج من أجهزة إدراك الكلام، وفي هذه المقالة نناقش استخدام شبكات عصبونية متعددة الطبقات من النوع Feed-Forward والتي يتم تدريبها باستخدام خوارزمية مرشح كالمن الثنائي المطور Dual Extended Kalman Filtering لإزالة اللاستقرارية والضجيج الملون من الكلام ومدى فعاليتها في هذا المجال، وقد بينت التجارب العملية المقدّمة فعالية هذه الإجرائية.

كلمات مفتاحية:
الشبكات العصبونية الاصطناعية – إشارة الضجيج – مرشح كالمن.

* أستاذ مساعد في قسم الاتصالات – كلية الهندسة الميكانيكية & الكهربائية – جامعة تشرين – اللاذقية –سوريا.
** مدرس في قسم الحاسبات والتحكم – كلية الهندسة الميكانيكية & الكهربائية – جامعة تشرين – اللاذقية –سوريا.
*** طالب ماجستير في قسم الاتصالات – كلية الهندسة الميكانيكية & الكهربائية – جامعة تشرين – اللاذقية–سوريا.

مجلة جامعة تشرين للدراسات و البحوث العلمية _ سلسلة العلوم الهندسية المجلد (28) العدد (2) 2006
Tishreen University Journal for Studies and Scientific Research- Engineering Science Series Vol. (28) No (2) 2006

Using Artificial Neural Networks for Removing Noise from Speech Signal

Dr. Lamis Kudsi *
Dr. Bilal Chiha **
Hussain Alazki***

(Accepted 10/10/2006)

□ ABSTRACT □

Improvement of quality and intelligibility of the speech signal by removing its noise has a wide range of applications; so it is necessary for improving the performance of vocal communications under noisy conditions such as telecommunications in aviation and teleconferencing and cellular communications.

The Artificial Neural Networks are one of the important developments aimed at removing noise from speech recognition devices. In this paper, we will also discuss the use of Feed-forward neural networks training by dual extended Kalman Filtering algorithm (DEKF) to remove nonstationary and colored noise from speech, presenting a number of experimental results supporting such a procedure.

Key Words: Artificial Neural Networks - Noise Signal - Kalman Filtering

*Assistant Professor, Communications Department, Faculty of Mechanical & Electrial Engineering, Tishreen University, Lattakia, Syria.
**Assistant Professor, Computer and Control Department, Faculty of Mechanical & Electrial Engineering, Tishreen University, Lattakia, Syria.
***Postgraduate Student, Communications Department, Faculty of Mechanical & Electrial Engineering, Tishreen University, Lattakia, Syria.

مقدّمة:

أحد أهم المشاكل الرئيسية لتقليل الضجيج من الكلام تكمن في صعوبة صنع نموذج مضبوط للكلام، وتعد الشبكات العصبونية الاصطناعية أحد التطورات المهمة في الترشيح اللاخطي لمجال الزمن حيث إن مجال تقانات تحسين الكلام التقليدية (الطرح الطيفي، التقريبات التكرارية لمجال الزمن.....إلخ)[1]. كل واحدة من هذه التطبيقات هي نتيجة تكرارية بتشويه سمعي لإشارة الكلام. ولهذا السبب الشبكات العصبونية تعتمد الترشيح باستخدام مجموعة معطيات من الكلام النقي وضجيج الكلام الاصطناعي لنمذجة الكلام المحسن برمجياً من ضجيج الكلام، حيث إن مجموعة معطيات من الكلام النقي تكون مقدّمة كإشارة هدف للتدريب. هذه الطرق بالغالب فعالة من خلال مجموعة التدريب، ولكن تميل لتعميم ضعيف لمنابع فعلية مع مستويات ضجيج وكلام متباينة (معلومات عن شبكات عصبونية تعتمد التقريب يمكن إيجادها في [13]). هذه الشبكات التي تعتمد تلك الطرق لا تكون ذات طبيعة مهمة نحو تأكيد الطبيعة اللااستقرارية.

في هذه المقالة استخدمت خوارزمية مرشح كالمن المطور الثنائي لتدريب شبكات عصبية متعددة الطبقات من النوع Feed-Forward المؤلفة من ثلاث طبقات كل طبقة مؤلفة من وحدات معالجة ترتبط بكل الوحدات الأخرى في الطبقة التالية الأعلى ولا تتصل فيما بينها ضمن الطبقة الواحدة. وفي هذه الخوارزمية التقريبية المقدّمة نفرض وجود فقط إشارة الضجيج والشبكات العصبونية تدرب على إشارة ضجيج كلام محدد.

خوارزمية مرشح كالمن المطور EKF طبقت أولاً لتدريب الشبكات العصبية المتعددة الطبقات Wu and Singhal [11]. وتبين أن خوارزمية EKF تتجه نحو طريقة أقل تكرارية من خوارزمية الانتشار العكسي back-propagation وهي تتجه لتكون حل جيد عندما تكون خوارزمية back-propagation ضعيفة.

الشبكة العصبونية من النوع Feed-Forward تعمل بنموذجين: التدريب والتذكر، حيث إن أوزان الشبكة تثبت بنموذج التذكر وفي نموذج التدريب الأوزان تتجدد باستعمال مرشح كالمن الثنائي المطور، وميزة هذه الشبكات هي بإيجاد نماذج للمشكلة بحيث إن معطيات الدخل لا تكون منفصلة عن مجموعة التدريب.

خوارزمية مرشح كالمن الثنائي المطور تقلل عدد خطوات الزمن المتطلبة للتدريب بسرعة كبيرة لهذا النوع من المشاكل مقارنة بالخوارزميات الأخرى(back-propagation) ومتطلبات حساب هذه الخوارزمية صعبة لكنها تعطي حجم شبكة أقل.

أهمية البحث وأهدافه:

تتجلى أهمية البحث في المزايا العديدة التي تقدمها الشبكات العصبونية الاصطناعية في إمكانية استنتاج نماذج لإشارات الصوت والضجيج بهدف تنقية إشارة الصوت من الضجيج، واستخدام تقنية خوارزمية كالمن التكرارية لتدريب الشبكة بوجود ضجيج ملّون أو أبيض عند الأنظمة اللاخطية والخطية. لكي تتلائم مع أي إشارة متكلم، ضجيج، أو مستوى ضجيج غير متوقع.

وبناءَ على ما تقدم فإن الهدف من هذا البحث هي القدرة على استخلاص الضجيج من إشارة الصوت للوصول إلى نسبة إشارة إلى الضجيج أكبر ما يمكن لتحسين أداء أجهزة الاتصالات الصوتية في الملاحة البحرية، الطيران العسكري، بيئة الخليوي. ومن ثم تقييم هذه الطريقة من أجل أنواع من الضجيج سواءَ كان ضجيج ملّون أو أبيض.

طريقة البحث:

اعتمد في هذا البحث طريقة المحاكاة الحاسوبية والنمذجة الرياضية لذا تم إتباع المنهجية الآتية:

- دراسة وتحليل الإشارات (الصوت& الضجيج) باستخدام توابع مكتوبة بلغة MATLAB .
- وضع الشبكة العصبونية ومن ثم تدريبها لنمذجة الصوت والضجيج.
- تنفيذ المحاكاة باستخدام الحاسوب.
- مناقشة نتائج الدراسة وصياغة الاستنتاجات.

1- نموذج الكلام:

بوجه عام يوصف ضجيج الكلام بالعلاقة الآتية:

$$y_k = h(x_k, n_k) \quad (1)$$

حيث أن x_k : إشارة الكلام النظيف المقاد بالضجيج.

$h(.)$: قناة الاتصال.

n_k : الضجيج المعالج.

تحليل اشارة الكلام يمثل بـ y_k والتي يمكن أن تكون نموذج دقيق للارتباط الذاتي اللاخطي مع كل من الضجيج الملاحظ الإضافي والمعالج:

$$x(k) = f(x(k-1),.....,x(k-M),w) + v(k) \quad (2)$$

$$y(k) = x(k) + n(k) \quad (3)$$

$v(k)$: الضجيج المعالج.

$n(k)$: الضجيج الملاحظ.

w : الأوزان.

M : عدد القياسات.

$f(.)$: تابع لاخطي من قيم سابقة للمتحول $x(k)$ التي لها أوزان. ويفترض أن يكون الكلام مستقرا فقط فوق قطاعات صغيرة وبكل قطاع يملك الكلام نموذج مختلف والملاحظة المتوفرة تكون $y(k)$: التي تحتوي على الضجيج الإضافي.

معظم الطرق المباشرة لتقريب التوقع الشرطي $E[x(k)/y(k)]$ (توقع $x(k)$ بمعرفة $y(k)$) تتطلب أن تدرب الشبكات على مجموعة من معطيات نظيفة $x(k)$ تستخدم كهدف للشبكات العصبونية. ولعدم توافر الكلام النظيف يكون الهدف هو تقدير $x(k)$ من قياسات الضجيج $y(k)$ فقط.

لذلك نستخدم نموذج شبكات عصبونية متعددة الطبقات من النوع Feed-Forword (FFNN) ونحسب التوقع الثنائي لكل من حالات $\hat{x}$ والوزن $\hat{w}$ المعتمدة على تقريب مرشح كالمن الموضح في الفقرة 2.

2- مرشح كالمن الثنائي المطور:

يستخدم مرشح كالمن لتقدير الإشارات ضمن إطار فراغ الحالة (state-space) لتدريب الشبكة العصبونية

المستخدمة FF. حيث يعتمد مرشح كالمن على تضمين التنبؤ السابق مع قياس جديد. العالم كونوريرال Connoreral [9] يقترح استخدام مرشح كالمن المطور مع شبكات عصبونية لتقدير الحالة فقط. العالمان Poskorios and Feldkamp [10] يقترحان لتقدير الوزن ضمن إطار فراغ الحالة تطبيق مرشح كالمن بشكل فعال للشبكة العصبونية. ثم وسعت هذه الأفكار لتتضمن تقدير كالمن الثنائي لكل من الحالات والأوزان لإيجاد الاحتمال الأعظمي بشكل أفضل لتنبؤ لاخطي بواسطة العالم Eric Wan [14]. اعتمد مع تطوير الدراسات في سياق معالجة إشارة الكلام لإزالة الضجيج منها.

لتطبيق مرشح كالمن المطور EKF يجب أولاً إيجاد معادلة الارتباط الذاتي من المعادلة (4) و (5) بشكل فراغ الحالة:

$$X(k) = F[X(k-1)] + Bv(k) \quad (4)$$

$$Y(k) = CX(k) + n(k) \quad (5)$$

$$B = \begin{bmatrix} 1 \\ 0 \\ . \\ . \\ 0 \end{bmatrix} \quad , \quad Y(k) = \begin{bmatrix} X(k) \\ X(k-1) \\ \\ \\ X(k-M+2) \end{bmatrix} \quad \text{حيث أن :} \quad (6)$$

$$F[X(k)] = \begin{bmatrix} F(X(k), - - - - -, X(k-M+1), w) \\ X(k) \\ \\ X(k-M+2) \end{bmatrix} \quad (7)$$

$$C = B^{\mathrm{T}}$$

حيث أن B^{T} هو منقول المصفوفةB.

M: العينات.

إذا كان النظام خطي عندها $F[X(k)]$ يأخذ الشكل $w^{\mathrm{T}}X(k)$ و $F[X(k)]$ يمكن أن يكتب بشكل $AX(k)$، حيث إن A هي مصفوفة بصيغة قابلة للتحكم. نفترض أوليا الضجيج المعالج $v(k)$ والضجيج الملاحظ $n(k)$ هما ضجيج أبيض مع معرفة التشتت لهما σ_v^2 و σ_n^2 على التوالي.

2-1 تقدير الحالة:

في النظام الخطي مرشح كالمن (KF) يمكن أن يكون قابلا للتطبيق لتقدير الحالات[6]. ومن أجل كل خطوة زمنية المرشح يحسب احتمال المربعات الأصغرية الخطية $\hat{X}(k)$ والتنبؤ $\hat{X}^{-}(k)$ فضلا عن خطأ التشتت $P_X(k)$ وتتبؤ خطأ التشتت $P_X^{-}(k)$. في النظام الخطي باستخدام إحصائيات غوصية [6] الاحتمالات هي احتمال المتوسط التربيعي الأصغري بدون معلومات سابقة عن X وهذا يقلل من تقدير الاحتمال الأعظمي.

لا يستطيع مرشح كالمن (KF) أن يكون تطبيق مباشر للشبكة العصبونية في النظام اللاخطي ولكنه يتطلب مصحح خطي للنظام اللاخطي. وهنا استخدام خوارزمية مرشح كالمن المطور EKF الموضحة بالمعادلات الآتية:

$$\hat{x}^{-}(k) = F[\hat{x}(k-1), w] \quad (8)$$

$$P_x^{-}(k) = A(K)P_x(k-1)A^T(k) + B\sigma_v^2 B^T \quad (9)$$

$$A(k) = \frac{\partial F[\hat{x}, w]}{\partial \hat{x}(k-1)} \quad (10)$$

حيث أن:

$$K(k) = P_x^{-}(k)C^T(CP_x^{-}(k)C^{\mathrm{T}} + \sigma_n^2)^{-1} \quad (11)$$

$$P_x(k) = (I - K(k)C)P_x^{-}(K) \quad (12)$$

$$\hat{X}(k) = \hat{X}^{-}(k) + K(k)(y(k) - C\hat{X}(k)) \quad (13)$$

حيث I: هو محدد مصفوفة انتقال الحالة.

من المعادلة (10) نجد أن المشتق يساوي إلى المصحح الخطي للشبكة العصبونية وهذا يمكن إيجاده بتطبيق وحيد لشبكة عصبونية ذات انتشار عكسي قياسي (Bp). وعندما تكون الأوزان w غير متاحة لنا يجب التبديل بالاحتمال $\hat{w}$.

2-2 تقدير الأوزان:

بما أن نموذج الكلام غير معروف فإن خوارزمية مرشح كالمن المطور EKF القياسية لا يمكن أن تكون تطبيق مباشر [11]. وجد أن حل المسألة ببناء صيغة فراغ الحالة منفصل للأوزان على الشكل الآتي:

$$w(k) = w(k-1) \quad (14)$$

$$y(k) = F(X(k-1), w(k)) + v(k) + n(k) \quad (15)$$

حيث الانتقال من حالة إلى أخرى يرتبط بمحدد المصفوفة، والشبكة العصبونية $(F(X(k-1), w(k))$ تلعب دور ملاحظة لاخطية خلال زمن متغير للمتحول w. وتسمح معادلات فراغ الحالة بتقدير الأوزان مع مرشح كالمن مطور EKF آخر:

$$\hat{w}^{-}(k) = \hat{w}(k-1) \quad (16)$$

$$\hat{P}_w^{-}(k) = P_w^{-}(k-1) \quad (17)$$

$$K_w(k) = P_w^{-}(k)H^{\mathrm{T}}(k)[H(k)P_w^{-}(k) + \sigma_n^2 + \sigma_v^2]^{-1} \quad (18)$$

$$P_w(k) = (I - K_w(k)H(k))P_w^{-}(k) \quad (19)$$

$$H(k) = \frac{\partial X^{-}(k)}{\partial \hat{w}} \quad (20)$$

حيث أن:

$$\hat{w}(k) = \hat{w}^{-}(k) + K_w(k)(y(k) - X^{-}(k)) \quad (21)$$

K(k): ربح كالمن.

إن استخدام مرشح كالمن المطور EKF من أجل تقدير الحالة يتعلق بالمربع الأصغري التكراري (RLS) وهكذا يمثل من جديد طريقة ثانية أكثر ملائمة. وجد أنه عندما يكون X غير متوافر يجب تبديل مرشح الوزن بالاحتمال $\hat{X}$ ، وجد أن الترجمة الفورية للاحتمال الأعظمي لمرشح كالمن مطور EKF تتضمن استخدام X مقابل $\hat{X}$ [8].

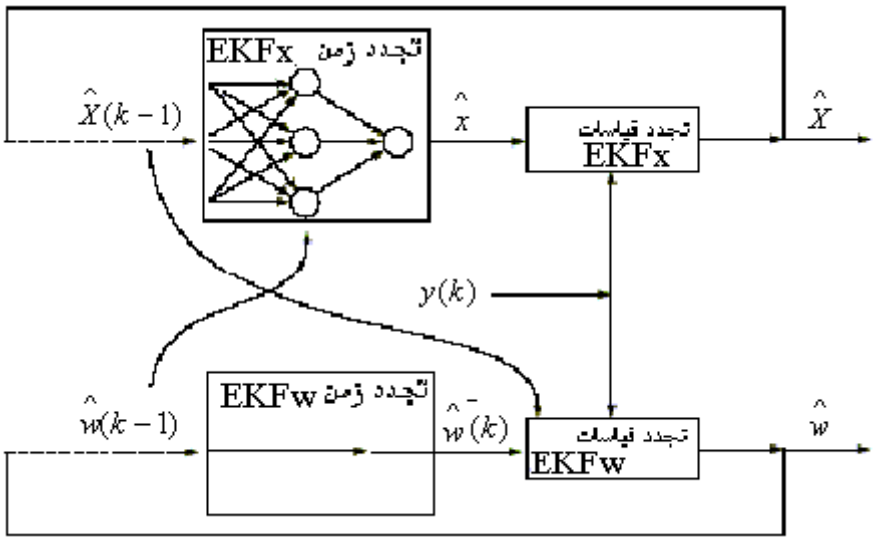
المصحح الخطي للشبكة العصبونية في المعادلة (20) يحسب بمشتق ديناميكي [14] ليفسر الطبيعة التكرارية لمرشح تقدير الحالة والمتضمن الاعتماد على ربح كالمن K(k) للأوزان.

حساب هذه المشتقات يكون صعب الحساب ويمكن تجنب ذلك بفضل اعتماد التوقع $\hat{X}(k-1)$ عن $\hat{w}$ في المعادلة (8) باستخدام المصحح الخطي المستقر للشبكة.

2-3 التقدير الثنائي للحالة والوزن:

تعتمد خوارزمية مرشح كالمن الثنائي المطور DEKF على تطبيق مرشح كالمن مطور للحالة و مرشح كالمن مطور للوزن على التوالي الشكل (1) لتجديد التقدير للمتحول $X(k)$ و w من أجل كل إزاحة زمنية.

من أجل كل خطوة زمنية التقدير الحالي للمتحول $X(k)$ يكون بمرشح وزن والتقدير الحالي للوزن w يكون بمرشح الحالة ومن أجل مجموعة معطيات محددة تطبق الخوارزمية بشكل تكراري إلى أن نصل إلى أصغر خطأ متوسط تربيعي.



الشكل (1) يوضح مرشح كالمن الثنائي المطور.

هذا التقريب للتقدير الثنائي يمكن أن يكون غير محدد ضمن إطار عمل الاحتمال الأعظمي ويمكن أن يكون أيضا متعلق بالتوقع الأعظمي EM [8] في حالة الخطية. وأيضاً [11] بتقريب عصبي لخوارزمية خطأ التنبؤ التكراري لإشارة الكلام هذه الطريقة متعلقة بشكل تقريبي بتقريب Lim [7] ليناسب نماذج LPC (ترميز التنبوء الخطي) للكلام المختلط بالضجيج، في هذا البحث نستعمل التقدير اللاخطي ليناسب البيانات المعطاة بدلاً من استعمال عدد

محدد من نماذج خطية مسبقة التحديد.

3- الشبكات العصبونية من النوع Feed –Forward:

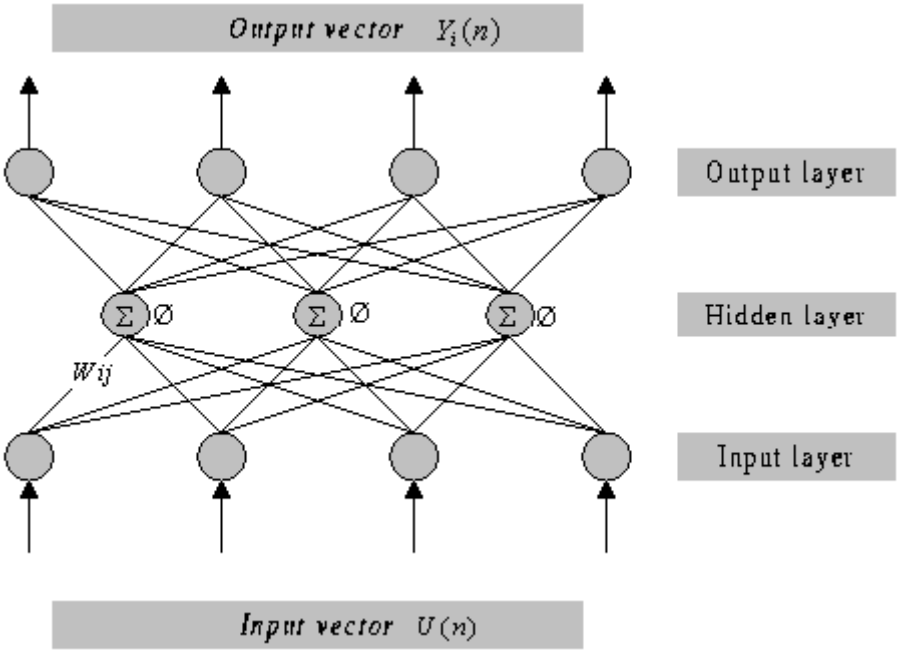
شبكات FF هي تقنية متكيفة قوية لعلاقات تقريبية بين قيم المداخل المستمرة والمخارج وهي عبارة عن إطار لتمثيل نموذج تحكمي. هذه الشبكة لها عدد من الأوزان n_w ورقم محدد من العقد n_n والتي فيها n_i هي عقد الدخل و n_o هي عقد خرج. العقد تكون مرتبة من دخل إلى خرج ومرتبة من 1 إلى n_n عندئذ العملية للشبكة يمكن أن توصف كالآتي:

$$Y_i(n) = \begin{cases} U_i(n) & for\ i = 1..... \quad n_j \\ \Phi\left(\sum_{j=1}^{i-1} WijYi\right) & for \quad i = n_i + 1.....n_n \end{cases} \qquad (22)$$

حيث $Y_i(n)$: هي مخارج العقد n_n للشبكة العصبية و Wij هي الربط الوزني لعقدة i إلى عقدة j.

$U(n)$: هي المداخل للشبكة n_i.

Ø : هي التابع الفعال (التنشيط) المستخدم.



الشكل (2) يوضح شبكة عصبونيةFF متعددة الطبقات.

هذا وصف لشبكة FF والمتضمنة طبقات من العقد يمكن إنتاجها بأوزان صفرية.

مخارج الشبكة n_o يمكن أن تكتب الآن كالآتي:

$$h_n(w(n),U(n)) = [h_1, h_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots h_{no}] = [Yn_n - n_o + 1, Yn_n - n_o + 2, \ldots \ , Yn_n] \quad (23)$$

حيث w هي $n_w \times 1$ شعاع يتضمن أوزان مرتبة w_{ij} للشبكة ولهذا السبب K_n (ربح كالمن المرتبط بالدليل n) يكون عبارة عن مصفوفة مشتقات $n_w \times n_o$ مستقرة لمخارج شبكة FF مع الأخذ بالحسبان أوزان الشبكة. من أجل شبكات Ff خوارزمية (Bp) الانتشار العكسي تستخدم لحساب المشتقات اللازمة.

يكون تدريب الشبكة ضمن إطار عمل خوارزمية مرشح كالمن الثنائي المطور بملاحظة المعادلات اللاخطية التي تحدد عملية الشبكة، حيث إن إيجاد الأوزان والحالات يكون باستخدام خوارزمية تكرارية (مرشح كالمن الثنائي المطور) الموضحة في الفقرة 2 بملاحظة معادلات مرشح كالمن.

عند تطبيق خوارزمية مرشح كالمن المطور EKF لتدريب الشبكة العصبونية فإن أوزان الشبكة تنظم كشعاع $w(n)$ وتدرب كشعاع حالة $x(n)$ للنظام. خوارزمية EKF تجدد الحالة بكل إزاحة زمنية وأيضا تجدد مصفوفة ارتباط خطأ الحالة. وتعمل الشبكة على إيجاد الأوزان تبعاً للاختلاف بين خرج الشبكة والهدف لتصغير متوسط الخطأ التربيعي.

تصحيح الأوزان يكون باستخدام تابع تنشيط Ø(.)=Tanh(.) (حيث أن الأوزان تمثل شعاع غير معروف تعلم برمجياً من معطيات الضجيج)، بارمترات مرشح كالمن تتجدد بكل دور بحيث إن تقدير الاحتمال الأعظمي وبارمترات النموذج هي التي تعطينا نموذج الضجيج.

4- التجارب والمناقشة:

لتحسين إشارة الكلام المضاف إليها ضجيج أبيض غير استقراري باستخدام الشبكات العصبونية الاصطناعية المهيئة On-Line ضمن إطار عمل خوارزمية مرشح كالمن الثنائي المطور DEKF لتدريب بارمترات الشبكة. نطبق نافذة إشارة Hamming 64[ms] (512 نقطة بتردد تقطيع 8 KHZ) ومع نافذة Hamming 64[ms] جديدة تبدأ كل 8ms (64 Point) والتي تصغر حجم الشبكة برمجياً.

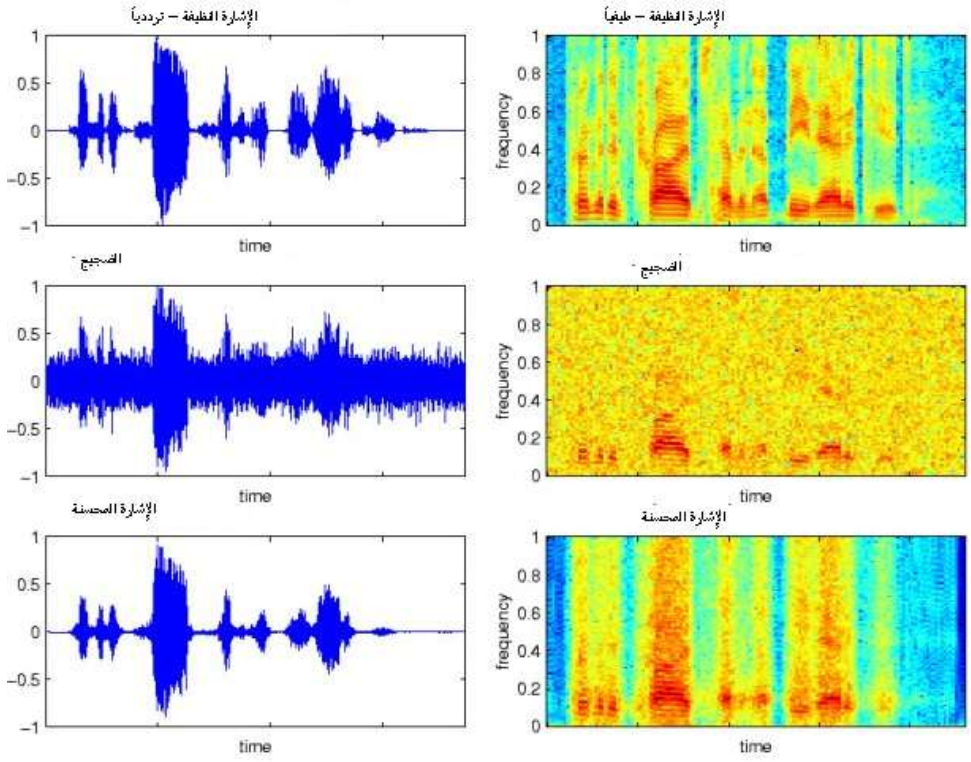
بعد معالجة الكلام وضجيج الكلام باستخدام توابع تحليل الكلام المكتوبة بلغة البرمجة Matlab قبل تطبيق المعطيات على دخل الشبكة المقترحة.

استخدام شبكة من النوع Feed-Forward متعددة الطبقات ثلاث طبقات مع 10 مداخل وأربعة وحدات في الطبقة المخفية وخرج وحيد مستعمل مع استخدام تابع تنشيط Ø(.)=Tanh(.). وبإدخال المعطيات إلى الشبكة بانزلاق نافذة المدخل كل 8ms. حيث إن أوزان الشبكة تجمع إلى شعاع وحيد الذي يتجدد حسب معادلات خوارزمية مرشح كالمن الثنائي المطور (المعادلات (8) .....(21)) ويكون الغاية من تدريب الشبكة هو إيجاد الأوزان تبعاً للاختلاف بين خرج الشبكة وإشارة الكلام لتصغير متوسط الخطأ التربيعي.

بارمترات النموذج تعلم (تعديل الأوزان والانحرافات) برمجياً من معطيات الضجيج لكل نافذة كلام لتسمح بتقدير متزامن لنموذج الكلام وإشارة الكلام حيث أن تقدير الاحتمال الأعظمي وبارمترات النموذج هي التي تعطي ماهية الضجيج باستخدام زمن ربح كالمن K_k ليقدم جودة أفضل مع تجدد الزمن من المعادلات (9) و (10) لإجراء التنبؤ الأولي من قيمة تالية من الحالة بشكل متتالي لنمذجة التنبؤ مع إجراء تقدير فراغ الحالة لترشيح مجال الزمن.

بعد ذلك نجد أنه عند تطبيق خوارزمية DEKF لإشارة الكلام من معطيات أساسية (عينات مأخوذة من منابع تقدير تحسين الكلام) [12] المضاف إليها ضجيج رشقي غير مستقر كما هو في الشكل (3) تحسب بافتراض أن كل من تشتت الإشارة σ_x^2 وتشتت الضجيج σ_n^2 مأخوذ بعين الاعتبار .متوسط نسبة الإشارة للضجيج SNR يكون 9,94[db].

وعندما σ_n^2 و σ_x^2 يتم حسابهم نستعمل إشارة الضجيج ومتوسط SNR يتحسن بمقدار 8,50[db] وبالمقارنة مع حالة الطرح الطيفي يتحقق تحسن 1,26[db].



الشكل (3) تحسن إشارة الكلام المضاف إليها ضجيج رشقي مع SNR=0 dB بدائية وتحسن بنسبة SNR=9.94 dB.

<u>الضجيج الملون:</u>

من أجل تدريب الشبكة العصبونية من النوع FF عند وجود الضجيج الملون نستبدل معادلات فراغ الحالة (3) و(4) قبل تطبيق تقنية مرشح كالمن، وذلك بقياسات الضجيج المعالج حيث تعطى معادلات فراغ الحالة:

$$N(k) = A_n N(k-1) + B_n v_n(k) \quad (24)$$

$$n(k) = C_n N(k) \quad (25)$$

حيث $N(k)$: شعاع من قيم مختلفة لـ $n(k)$ و ويكون عبارة عن ضجيج أبيض.

A_n : مصفوفة انتقال الحالة بصيغة قابلة للتحكم.

B_n و C_n : هي الشكل لـ Cو B نفسه المعطاة بالمعادلة (5).

هذه المساواة لنموذج الارتباط الذاتي للضجيج الملون ممكن أن تناسب مقطع صغير لإشارة الضجيج الملون

الذي لا يحتوي على كلام.

ومع هذه الصيغة للضجيج الملون يكون سهلاً التزايد لكل من الحالة X(k) والوزن W(k) مع n(k). وبالتالي المعادلات (3) و (4) تصبح :

$$\begin{bmatrix} x(k) \\ N(k) \end{bmatrix} = \begin{bmatrix} F[x(k-1)] \\ A_n N(k-1) \end{bmatrix} + \begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} v(k) \\ v_n(k) \end{bmatrix} \quad (26)$$

$$y(k) = \begin{bmatrix} CC_n \end{bmatrix} \begin{bmatrix} x(k) \\ N(k) \end{bmatrix} \quad (27)$$

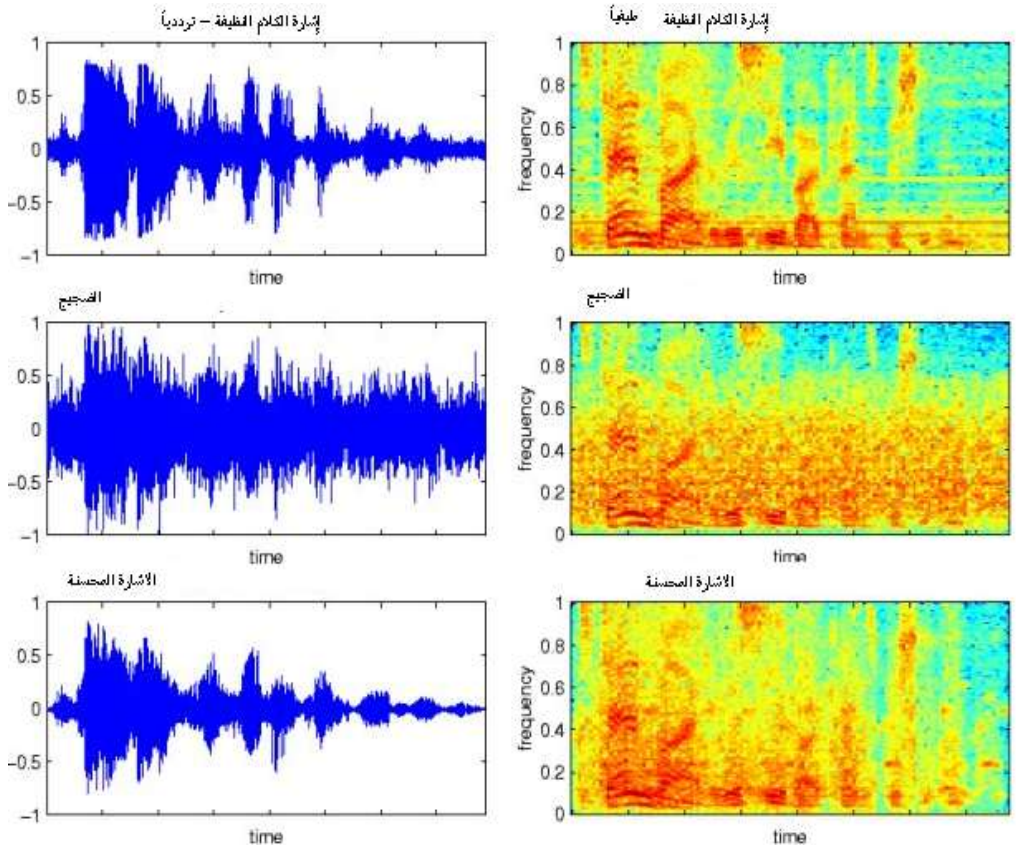
والمعادلات (12) و (13) تصبح على النحو الآتي:

$$\begin{bmatrix} w(k) \\ N(k) \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & A_n \end{bmatrix} + \begin{bmatrix} w(k-1) \\ N(k-1) \end{bmatrix} + \begin{bmatrix} 0 \\ B_n \end{bmatrix} v_n(k) \quad (28)$$

$$y(k) = f(x(k-1), w(k)) + C_n N(k) + v(k) \quad (29)$$

ويصبح الضجيج المعالج لمعادلات الحالة ضجيج أبيض وعندئذ خوارزمية DEKF يمكن أن تطبق على الضجيج الأبيض لتقدير إشارة الكلام. حيث أن الضجيج الملون يؤثر في تقدير الحالة وأيضاً بتقدير الوزن.

بإدخال معطيات للشبكة بانزلاق نافذة الدخل وهي ضجيج في هاتف خليوي مضاف إلى إشارة الكلام كما في الشكل (4) مع اختيار نسبة إشارة الضجيج بدائية وهي [db]0,16- حيث بتدريب البارامترات حسب معادلات خوارزمية مرشح كالمن المطور الثنائي مع شبكة عصبونية FF (10 مداخل و 4 وحدات مخفية وخرج وحيد). تعطي النتائج نسبة إشارة إلى الضجيج وهي SNR=5,60[db].



الشكل (4) تحسن إشارة الخليوي المضاف إليها ضجيج رشقي مع SNR=-0.16 dB بدائية وتحسن بنسبة SNR=5.60 dB.

باستخدام معطيات إضافية [12] مع ضجيج Pink مضاف إلى إشارة الكلام مع الافتراض أنه في كل حالات تجارب الضجيج الملون فإن نموذج الضجيج A_n وتشتت الضجيج المعالج σ_n^2 و σ_x^2 هي مقادير محسوبة حيث تتم النمذجة بإدراك مسبق لـ $x(k)$ و $n(k)$ وتدرب هذه الشبكة مع قيم مقدرة لضجيج الكلام فقط. والجدول(1) يلخص النتائج لمستويات ضجيج SNR مختلفة مع مقارنة مع تقانة الطرح الطيفي.

الجدول (1) يوضح مقارنة بين الطرح الطيفي و شبكة FF مبنية على خوارزمية DEKF.

ضجيج خليوي	الطرح الطيفي	DEKF
SNR ابتدائي [db]	-0,16	-0,16
SNR نهائي	2,48	5,60
SNR ابتدائي [db]	10	10
SNR نهائي	3.54	13.99
ضجيج Pink	***الطرح الطيفي***	***DEKF***
SNR ابتدائي [db]	5	5
SNR نهائي	3,24	8,88
SNR ابتدائي [db]	10	10
SNR نهائي	3.56	13.87

الاستنتاجات والتوصيات:

- الشبكات العصبونية الاصطناعية تقترب من التوقع الشرطي عندما جميع الإشارات الإحصائية تكون غوصية (نقاوة غير واقعية لإشارات الكلام ومنابع ضجيجية عالمية حقيقية)، وهذا يكون دافع قوي لاستخدام الشبكات العصبونية الغير خطية.
- استخدام طرائق forward-backward في الشبكات العصبونية الاصطناعية من النوع feed-forward تلعب دور في النمذجة اللاخطية لإشارة الكلام حيث أن معطيات الدخل لا تنفصل عن مجموعة التدريب.
- خوارزمية مرشح كالمن الثنائي المطور تقلل عدد خطوات الزمن المتطلبة لتدريب الشبكة العصبية الاصطناعية، ويبدو ذلك في تقليل حجم الشبكة (10 مداخل 4 وحدات مخفية وخرج وحيد).
- تدريب أوزان الشبكة العصبية الاصطناعية ضمن مجال خوارزمية مرشح كالمن المطور الثنائي لمعالجة الضجيج الأبيض اللاستقراري تعطي نسبة الإشارة إلى الضجيج [db]9,9 SNR.
- في معالجة الضجيج الملون لتحسين إشارة الكلام باستخدام شبكات FF المدربة باستخدام خوارزمية مرشح كالمن الثنائي المطور تعطي [db]5,60 SNR.
- استخدام تقنية التقريب التكراري في خوارزمية DEKF لاقتفاء أثر الضجيج اللاستقراري بملاحظة العلاقة بين خطأ التنبؤ المتبقي لضجيج الكلام والكلام النظيف تعطي SNR 5,60
- التقدير الدقيق لإحصائيات الضجيج يبقى مرتبط بالخوارزميات التي تستخدم لتلك التقديرات بحيث تتحسن هذه التقديرات أو تستخدم تقانات جديدة متطورة تلغي الحاجة لمعرفة واضحة بإحصائيات الضجيج.
- إزالة الضجيج من الكلام يبقى مسألة تحدي لإعطاء استنتاجات واضحة والشبكات العصبونية تقدم أداة فعالة لذلك عن طريق استخدام شبكات أكثر تطوراً لحساب المشتقات اللازمة.

المراجع:

1- DELLER, S.and PROAOKIS, S.- *Discrete-Time Prossing of Speech Signals*, Macmillan Publishing Company ,NY, 1993, 800p.

2- JULIER, S.J. and UHLMAN, J.K.- *A General Method for Approximating Nonlinear Transformations of Probability Distributions*, Technical report, RRG, dept, of engineering science, university of oxford .1996 ,www.robots.ox.ac.uk.

3- LEWIS, F.- *Optimal Estimation*, Sohn Wiley & Sons, Inc. New York, 1986, 358p.

4- Matthews, M. B. and Moschytz, G. S.- *Neural-network nonlinear adaptive filtering using Extended Kalman Filter Algorithm*, International Neural Network Conference, Paris, 1995, 115-118p.

5- MA, N. and BOUCHARD, M. and GOUBRAN, R.- *Perceptual Kalman Filtering for Speech Enhancement in Colored Noise*, Proc. Of IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP), volume: 1, 2004, 717-720p.

6- NELSON, A.T. and WAN, E.A.- *A Two-Observation Kalman Framework for*

Maximum-Likelihood Modeling of Noisy Time Series, Oregon Graduate Inst, volume: 3, 1999, 2489-2494p.

7- NELSON, A.T., and Wan, E.A.- *Neural Speech Enhancement Using Dual Extended Kalman Filtering*, in Proc, ICNN'97, Houston, TX, 1997, 2171-2175p.

8- NELSON, L. and STEAR, E.- *The Simultaneous On-Line Estimation of Parameters and States in Linear Systems*, Volume: 21, 1976, 94- 98p.

9- ONNOR, S. MARTIN, R. ATLAS, L.- *Recurrent Neural Networks and Robost Time Series Prediction*, IEEE Transactions on Neural Networks, Volume: 5, 1994, 240-254p.

10- PUSKORIOUS, G. FELDKAMP L.- *Neurocontrol of Nonlinear Dynamics Systems with Kalman Filter Trained Recurrent Networks*, Volume: 5, 1995, 279-297p

11- SINGHAL, S., WU, L.- *Training Multilayer Perceptrons with The Extended Kalman Algorithm, Advances in Neural Information Processing Systems***,** (Morgan Kaufmann, San Mateo), 1989, 133-140p.

12- SPAAR: *Speech Enhancement Assessment Resource*, Oregon Graduate Institute, http://ece.ogi.edu/NESEL/data.

13- WAN, E. A. and NELSON, A.- *Networks for Speech Enhancement*, in Handbook of Neural Networks for Speech Processing, Edited by Shigem Katagiti, Attech House, Boston, (First Edition), 1998.

14- WAN, E. A. and VAN, R.- *Kalman Filtering and Neural Networks*, Ed. Simon Haykin,. Wiley, 2002, 300p.