مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (6) العدد(2) 2022

Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (6) No. (2) 2022

طريقة حل مشكلة الترميز التنبؤي الخطي اعتماداً على ضغط الكلام

د. غسان محمد محمد*

(تاريخ الإيداع 3 /8 /2021 . قُبِل للنشر في 6 /3 /2022)

□ ملخّص□

يلقي هذا البحث الضوء على انواع ترميز الكلام واهمية ضغطه اقتصاديا وعمليا مع التركيز بشكل خاص على الترميز التنبؤي الخطي والذي غالبا ما يستخدم في معالجة الإشارة الصوتية ومعالجة الكلام، لتمثيل الغلاف الطيفي للإشارة الرقمية للكلام في شكلها المضغوط، وذلك باستخدام معلومات من نموذج تنبؤي خطي باعتباره من أفضل تقنيات تحليل الكلام، والطريقة الأكثر استخدامًا في ترميز وتوليف الكلام، وإحدى الطرق الأكثر فائدة لترميز كلام ذي نوعية جيدة في معدل بت منخفض.

تم محاكاة عملية الترميز التنبؤي الخطي من خلال برنامج الماتلاب، وتم التوصل الى نتائج وتوصيات تساعدنا في فهم اكبر لهذه التقنية وسبب استخدامها الكبير.

كلمات مفتاحية: الترميز التنبؤي الخطي ، LPC ، ترميز الكلام ، نموذج تنبؤي خطي

*أستاذ مساعد – قسم هندسة تكنولوجيا الاتصالات – كلية هندسة تكنولوجيا الاتصالات و المعلومات – جامعة طرطوس .

مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (6) العدد(2) 2022

Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (6) No. (2) 2022

A study of linear predictive codingbased on speech pressure

Dr.Eng.Ghassan Mohammed Mohammed*

(Received 3/ 8/ 2021 . Accepted 6/ 3/ 2022)

□ ABSTRACT □

This research sheds light on the types of speech coding and the importance of its compression, economically and practically, with a special focus on linear predictive coding, which is often used in audio signal processing and speech processing, to represent the spectral envelope of the digital speech signal in its compressed form, using information from a linear predictive model which is One of the most effective speech analysis techniques, the most widely used method for speech encoding and speech synthesis, and one of the most useful methods for encoding good quality speech at a low bit rate, it provides highly accurate estimates of speech parameters.

The linear predictive coding process in all its stages has been simulated by Simulink Matlab, and many results and conclusions have been reached that help us in a greater understanding of this technology and the reason for its great use.

Key Words: Linear predictive coding, LPC, speech coding, linear predictive model, LPC simulation**.**

*Associate Professor, Department of Communications Engineering, Faculty of ICT Engineering, Tartous, University, Tartous, Syria.

1- مقدمة:

يتسم القرن الواحد والعشرون بتسارع كبير في وتيرة التطور التكنولوجي، حيث يشهد العالم بأكمله تطوركبير في مجالات التعرف على الكلام وضغطه ونقله وترميزه و إنتاجه ، في ضوء هذا كله كان هناك حاجة كبيرة لاستخدام أفضل طرق الترميز الفعال لتحقيق ضغط جيد للكلام المنقول والحصول على كلام أوضح و أصفى واقل تشويش من الكلام الأصلي بالإضافة لفهم أفضل لعملية مراحل توليد ومعالجة الكلام.

2- هدف البحث وأهميته:

الغاية الأساسية من هذا البحث التعرف على الترميز التنبؤي الخطي وأهميته الكبيرة في ضغط المعلومات الطيفية بما ينعكس على عملية نقل الكلام من الناحية الاقتصادية و نمذجة الصوت عن طريق دراسة مشاكل الترميز التنبؤي الخطي ومحاولة إيجاد انسب الحلول رياضيا وهذا بدوره يؤدي إلى فرض قيود على استخدام الترميز التنبؤي الخطي في هذا البحث ، وأهمية هذا البحث تكون في الاستخدامات الواسعة للترميز التنبؤي الخطي على سبيل المثال في نظام الاتصال النقال(Global System Mobile)GSMحيث يتم استخدامه في ضغط الصوت مما يمكن من إرساله عبر قناة صوتيه ضيقة وهذا مفيد من الناحية الاقتصادية، ,وأيضا استخدامه لبناء مشفر صوتي(coder)، والمستخدم كإشارة تحفيز لمرشح متغير في الزمن الحقيقي وخاصة في الموسيقى الالكترونية [1].

3- طريقة البحث:

اعتمدنا في هذا البحث منهجية المحاكاة الحاسوبية وذلك من خلال التعرف على العلاقة بين الصوت والكلام وعملية ضغط الكلام بالإضافة إلى شرح مختصر عن الموجات الصوتية والتعرف على مراحل نظم التعرف على الكلام وانواع ترميز الكلام والتركيز على مشاكل الترميز التنبؤي الخطي وإيجاد حلول لها وما هي القيود التي يفرضها عند استخدامه وذلك من خلال برنامج الماتلاب.

تم مناقشة النتائج واستخلاص الاستنتاجات والتفسيرات بالإضافة إلى قائمة بأسماء المراجع التي يستند اليها البحث.

4- الموجات الصوتية منشؤها وأنواعها:

هي عبارة عن حركة الطاقة التي تنتقل عبر الوسط الناتجة عن اضطراب بعض المواد كالماء والهواء أو أي مادة سائلة أو غازية والأمواج الخاصة بالأوساط السائلة والغازية تسمى الموجات الطويلة والامواج الموجودة بالأوساط الصلبة تسمى بالموجات المستعرضة.

تتشكل الموجات الصوتية نتيجة لاهتزاز الهواء وتنقل هذه الموجات الصوتية عبر الهواء بنفس طريقة انتقال موجات الماء[4].

يوجد ثلاث انواع مختلفة للأمواج الصوتية [6]:

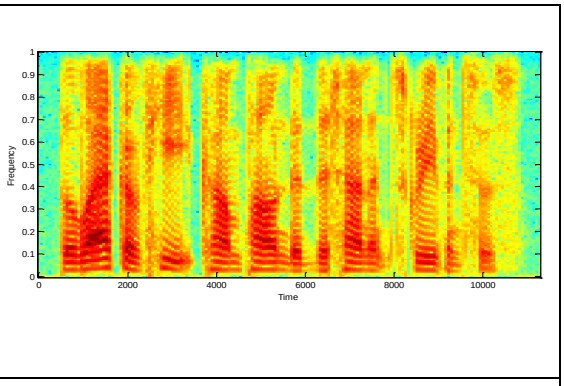
- الموجات السمعية: يستطيع الإنسان سماعها وادراكها وينحصر ترددها ضمن المجال الترددي من20Hz إلى 20KHz.
- الموجات فوق السمعية : لا يستطيع الانسان سماعها وادراكها حيث ترددها يزيد عن20KHz.
- الموجات دون السمعية: لا يستطيع الإنسان سماعها وإدراكها حيث يقل ترددها عن 20Hz.

4-1- مميزات الأمواج الصوتية:

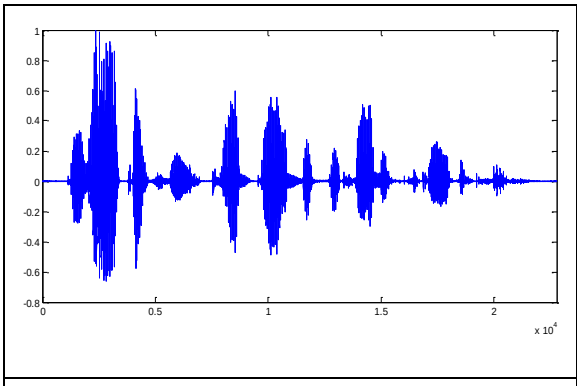
1. **ظاهرة انكسار ومرور الأمواج:** عندما تواجه موجة جيبيه متقدمة حاجزا أو فتحة صغيرة، يتغير انتشارها: مما يؤدي الى تشوهها. وهذا يسمى بظاهرة انكسار الموجة وتحدث هذه الظاهرة في حال كانت أبعاد الفتحة أو الحاجز من رتبة طول الموجة.
2. **الامتصاص:** يستخدم مفهوم امتصاص الموجات الصوتية لعزل الغرف من الصوت وتغطية الجدران بمواد لينة حيث ان الموجة الصوتية هي موجة ضغط وعندما تصطدم بحاجز لين ستتخامد الموجة ويصغر مطالها او اتساعها وبالتالي نقلها بشكل غير جيد.
3. **انعكاس الأمواج الصوتية:** تصطدمالموجة الصوتية بحاجز ذو أبعاد أكبر من طول موجتها، فتنعكس الموجة وبالتالي فان اتجاهات انتشار الصوت الوارد والمنعكس يشكلان مع الناظم زاويتين متساويتين.
4. **انحراف الموجات الصوتية** :هو قدرة الموجة لتجنب حاجز أو المرور عبر فتحة وبفضل الظاهرة نتمكن من سماع الصوت من وراء الحاجز.
5. **ظاهرة الصدى:** حيث ينعكس الصوت على حاجز جامد يملك ابعاد أكبر بكثير من طول موجة الصوت[1].

4-2- الصوت والكلام:

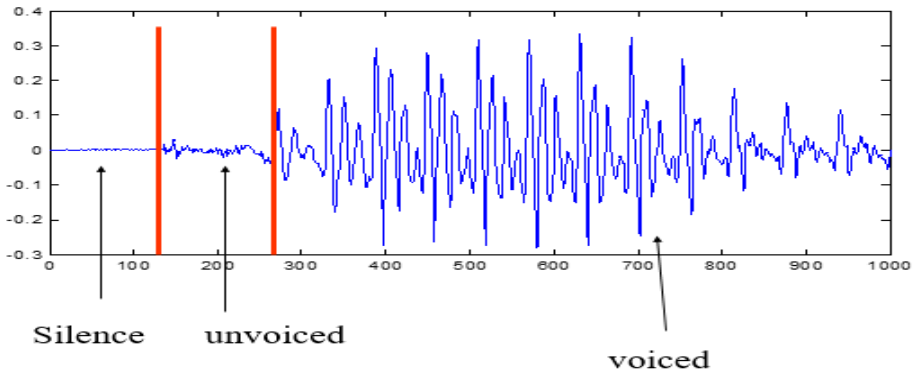
يوصف الكلام بأنه يمثل الصوت الناتج من جهاز التكلم الخاص بالإنسان لإنشاء تمثيل لغوي ينقلالمعلومات من المرسل الى المستقبل، ويتضمن مفهومه المحتويات الحقيقة لإشارة الكلام، معلومات عن شخصية المتكلم وهويته، الانفعال ، التنغيم ، اضافة الى الرسالة المنقولة بالكلام ، اما المعلومات المتضمنة في الاشارة الصوتية تتمثل بالشكل الفسيولوجي للجهاز الصوتي واللهجة والحالة النفسية ، حيث يبين الشكل (1) نموذج موجة الصوت والشكل (2) الطيف الترددي للموجة الصوتية والشكل (3) يوضح الفرق بين الصوت المسموع وبدون صوت وحالة الصمت [1]



الشكل (1) نموذج موجة الصوت



الشكل (2) الطيف الترددي للموجة الصوتية



الشكل (3) يمثل الفرق بين الصوت المسموع وبلا صوت والصمت

5- ضغط الكلام :

هو عملية تحويل لصيغة المعطيات لتحتوي على عددا اقل من البتات ، مما يقلل حجم البيانات المرسلة ، وبالتالي إمكانية تخزين وإرسال معطيات اكبر. يوجد نوعين من الضغط [2]:

- طرق الضغط بالضياع : التشفير المتكرر وطريقة هوفمان و طريقة LZW.
- طرق الضغط بدون ضياع : طريقة التحويل الموجي التي هي محور هذا البحث، وطريقة المكمم الاتجاهي وطريقة شفرة التنبؤ.

6- مراحل التعرف على الكلام:

اول مرحلة من نظم التعرف على الكلام هو تحويل الإشارة الصوتية إلى إشارة رقمية من خلال تقطيع (sampling)الاشارة الصوتية التي تم استقبالها ومن ثم تكميمها (quantization) وبعد عملية التكميم يتم ترميز مطالات العينات المأخوذة من الإشارة الأصلية باستخدام المرمز(Encoder) الذي يخزن القيم الثنائية إما *8bit* أو *16bit*.

المرحلة الثانية هي التعرف على الكلام و نمذجته إحصائياً من خلال أنظمة إحصائية متعددة بسبب صعوبة استيعاب الكلام المستمر والمتسلسل وبسبب تشابه بعض كلمات اللغة في الصوت وطريقة النطق، ومن أكثر هذه الأنظمة شيوعاً نموذج ماركوف المخفي (the hidden Markov model) والشبكات العصبونية (neural networks)[2].

6-1-بارمترات التعرف على الكلام:

اهم البارامترات(parameters) التي تميز نظم التعرف المختلفة هي:

1. طريقة لفظ الكلام: نميز بين نظم الكلمات المعزولة حيث يتوقف المتكلم برهة بين كلمة واخرى ،والكلمات المتصلة مثل ارقام الهاتف التي لا تطلب نفس التوقف، ونظم التعرف العفوي والارتجالي وهو اصعب من النظم السابقة .
2. حجم المفردات المختلفة التي يمكن أن يتعرف عليها النظام: صغير (حتى مئة مفردة) ، متوسط (يصل إلى 1000 مفردة)، واسع (يزيد على 1000 مفردة) .

3. علاقة النظام بالمتحدثين: يوجد ثلاث أنواع من نظم التعرف: نظم مرتبطة بالمتحدث، حيث تتطلب إدخال عينات من كلام كل مستخدم جديد، ونظم مستقلة عن المتحدث independent speaker لا تتطلب العملية السابقة، ونظم متكيفة مع المتحدث تعمل على تحسين أدائها مع كل متحدثأثناء استخدامه للنظام.
4. مقدار المعرفة الصوتية والمعجمية في نظم التعرف: تراوح من نظم بسيطة لا تتضمن أي معلومات لغوية إلى نظم معقدة متكاملة المعرفة الصوتية واللغوية النحوية والدلالية والسياقية.

6-2-صعوبات التعرف على الكلام وحلولها:

تكمن صعوبة مسألة التعرف على الكلام في التغيرات التي تطرأ على شكل الإشارة الصوتية الموافقة للكلمة نفسها، وذلك بسبب التغيرات التي يمكن تصنيفها وفق الآتي:

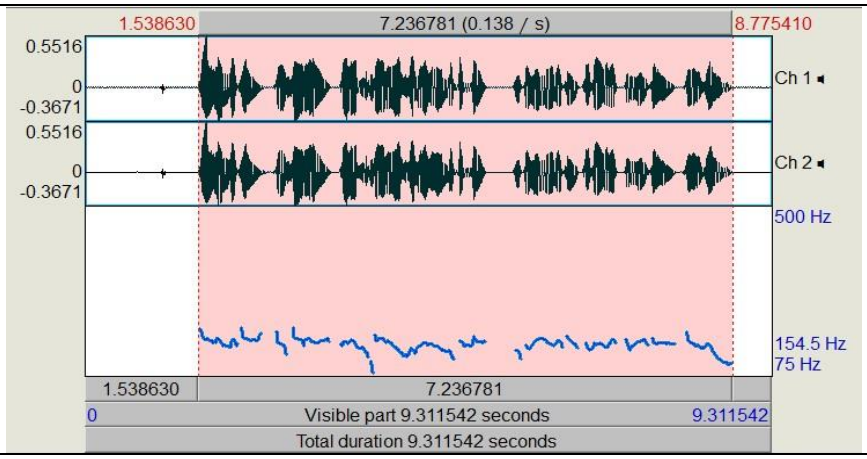
1. تغيرات تتعلق بالمتحدث نفسه: إذ يمكن أن تنشأ تغيرات صوتية بحسب تغيرات الحالة الفيزيولوجية والنفسية للشخص ، هل هو مريض أو بصحة جيدة ، حزين أو فرح أو غاضب...وبحسب معدل كلامه (سريع أو بطيء)، وبحسب جودة صوته(يقصد الإفهام أو يتكلم على نحو عارض).
2. تغيرات بين المتحدثين: تتعلق بالخلفية الاجتماعية اللغوية لهم مثل: اللهجة وأبعاد وشكل جهازهم الصوتي الفسيولوجي.
3. التحقيق الصوتي للصوتيات: وهي أصغر الوحدات الصوتية التي يمكن أن تتركب منها الكلمات، ويعتمد هذا التحقيق اعتمادا كبيرا على السياق الذي تلفظ فيه. على سبيل المثال، يختلف لفظ الصوت /ب/ من الكلمة (باب) إلى الكلمة (سبت)، فهو في الاولى /b /وفي الثانية / p /.
4. اختلاف محيط المتحدث: وهذا ناتج عن وجود إشارات صوتية غير مرغوب فيها (متحدثين آخرين، ضجيج، إغلاق باب...)، أو عن تنوع الميكرفونات المستخدمة ومكان توضعها.

6-3-سمات الصوت:

يوجد العديد من السمات الصوتية التي يجب أن تؤخذ بعين الاعتبار في نظم التعرف على الكلام:

1. كثافة الصوت (Intensity): وهي الطاقة التي تحملها موجات الصوت لكل وحدة من الوسط المحيط.وتقاس بالواط للمتر المربع (w/m2).

2. الحدة(pitch) : وهي الجودة التي تجعل من الممكن الحكم على الأصوات بأنها "أعلى" و "أقل "بالمعنى المرتبط بمستوى الصوت أو اللحن وهي تعتبر كمياً كتردد، ولكنها لا تعتبر كمية فيزيائية حيث لا تملك الموجات الصوتية حدّة ولكن يمكن قياس الذبذبات للحصول على التردد، ويعتبر تحديد الحدة الصوتية مهماً جداً للكثير من خوارزميات معالجة الكلام، ويوضح الشكل (4) الحدة لمتحدث ذات قيمة5.154Hz .

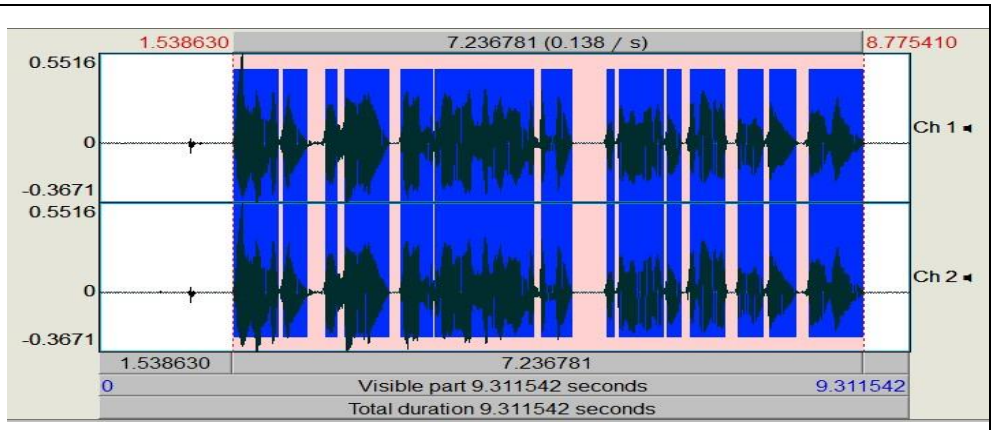


الشكل (4) تحديد الحدة الصوتية لمتحدث

3. طاقة النبضات (power of pulses): يُعتبر المطال للمقاطع اللاصوتية أقل بشكل ملحوظ من مطال المقاطع الصوتية. وتنعكس الطاقة قصيرة الأمد للإشارة الصوتية على شكل تغيرات في المطال (تذبذب في المطالات) وبالتالي فان طاقة الصوت هي شكل من أشكال الطاقة تقاس بالجول، فالصوت هو موجة ميكانيكية وتقاس كما هو مبين في الشكل (5) وفق القانون التالي[3]:

$$w = w_{pot} + w_{kin} = \int (\frac{p^2}{2\rho^0 c^2})dv + \int (\frac{\rho v^2}{2})dv \qquad (1)$$

حيث أن: P: سرعة الصوت، V: سرعة الجسيمات ، $\rho 0$: الكثافة ،C: سرعة الصوت.



الشكل (5) قياس طاقة النبضات

4. نسبة الاشارة الى الضجيج (Signal to noise ratio) SNR: تعتبر نسبة الإشارة إلى الضجيج SNRسمة هامة في تحديد نوعية البيانات السمعية، وتظهر أهمية هذه النسبة في تقنيات التعرف على الكلام وذلك لأن أداء التعرف على الأشخاص يتأثر بشكل كبير بها. لسوء الحظ في معظم التطبيقات لا يمكن تحديد نسبة الإشارة إلى الضجيج بدقة وذلك لأن طاقة الضجيج غير معلومة، يجب أن نحدد طريقة إحصائية تعطينا أفضل تقدير لنسبة الإشارة إلى الضجيج نظرا للمعرفة المسبقة لبيانات الكلام، و إحدى هذه الطرق هو تحليل إشارة الكلام قصيرة المدى لتوصيف النسبة SNR إحصائياً.

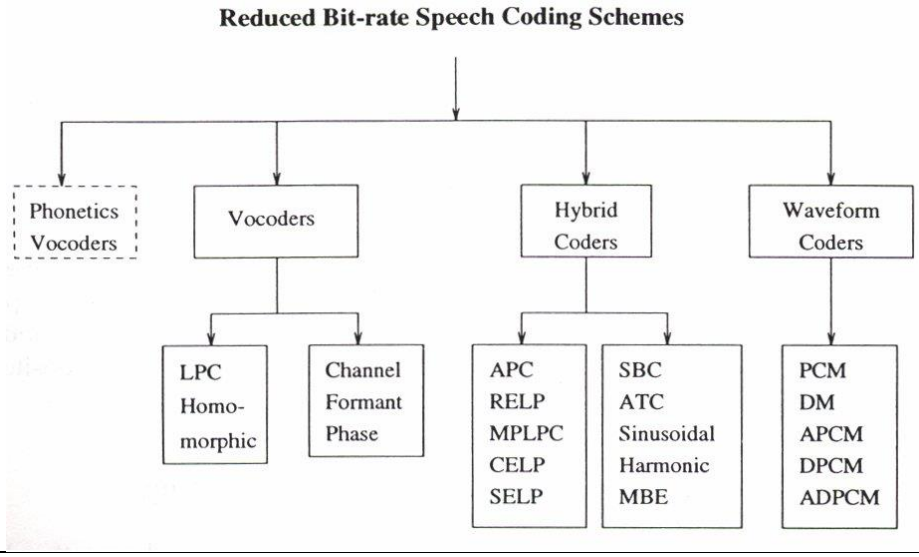
5. الجهارة (Shimmer): في العديد من التطبيقات التي تعتمدعلى القياسات الصوتية، فإن أجهزة الاستشعار النهائية هي الأذن البشرية، لهذا السبب عادة ما تحاول القياسات الصوتية وصف التصور الشخصي للصوت. تم بناء أجهزة لتوفير استجابة خطية، ولكن الأذن هي جهاز استشعار غير خطي، لذلك تم إيجاد مرشحات خاصة والمعروفة باسم مرشحات الترجيح والتي تستخدم للحسابات غير الخطية ،تتطلب المعايير العلمية المعتمدة أن تكون مقادير المرشح ضمن قناع التسامح المحدد، حيث تحدد المعايير اثنين من المجالات، مجال مع قيم التسامح الأكثر دقة ومجال مع قيم متسامح اقل دقة مثل مرشح من الفئة 1 ، بينما يشار إلى المرشح الذي يستوفي مواصفات التسامح للقناع الأكثر صرامة كفلتر من الفئة 1، بينما يشار إلى المرشح الذي يستوفي مواصفات المجال الأقل دقة بأنه مرشح من الفئة 2.

6. مقدار الارتعاش(Jitter): هو الانحراف عن الشكل الدوري المفترض للإشارة بالنسبة لإشارة ساعة مرجعية. في تطبيقات استعادة إشارة الساعة يسمى هذا انحراف توقيت الإرسال، وعند تصميم وصلات الاتصالات كون قيمة الانحراف كبيرة غير مرغوبة. من الممكن قياس انحراف الإرسال بنفس طرق قياس الإشارة المتغيرة زمنياً [3].

7-أنواع ترميز الكلام:

هناك ثلاث أنواع من ترميز الكلام[4]:

1. الترميز الموجي (Waveform coders) : يحاول هذا الترميز الحفاظ على الشكل العام لأشكال موجة الإشارة ويعمل بشكل عام على أساس عينة إلى عينة و عادةً ما يتم قياس أدائها بواسطة نسبة الإشارة إلى الضجيج SNR حيث أن التكميم هو المصدر الرئيسي للتشويه (الضجيج)، يتميز هذا النوع بجودة عالية في الترميز وتعقيد منخفض ومعدل بت مرتفع وتأخير منخفض.
2. الترميز الصوتي (Vocoders (voice coder: يتكون المرمز الصوتي من محلل ومركب. يستخرج المحلل من الكلام الأصلي مجموعة البارامترات التي تمثل نموذج توليد الكلام، والتي يتم إرسالها بعد ذلك، ثم يعيد المركب بناء الكلام الأصلي بناءً على البارامترات المرسلة وذلك بأخذ عينات من الكلام وتقسيمه إلى إطارات (حوالي 25 مللي ثانية)فبدلاً من نقل الكلام يتم بناء نموذج الكلام ومن ثم نقل بارامترات النموذج و تقريب الكلام المجمع ، تعد المرمزات (المشفرات) الصوتية خاصة جدًا بالكلام ويستهدف هذا المعيار بشكل أساسي التطبيقات الغير تجارية ، مثلا لأنظمة العسكرية الآمنة ، ومن خصائصه الجودة المنخفضة والتعقيد المعتدل ومعدل بت منخفض وتأخير مقبول.
3. الترميز المختلط (Hybrid coders): يجمع بين ميزات الترميز الصوتي والترميز الموجي ويتمثل بتشفير مجال التردد في تقسيم طيف اشارة الكلام إلى نطاقات أو مكونات ترددية باستخدام المرشح أو تحليل تحويل الكتلة. بعد التشفير وفك التشفير، تُستخدم هذه المكونات لإعادة تكوين شكل موجة الإدخال الأصلي بناءً على تجميع المرشح أو تحويل الكتلة العكسي. عادة يكون الهدف من تشفير المجال الزمني هو التنبؤ الخطي. ويمكن نمذجة الخصائص الإحصائية لإشارات الكلام بدقة شديدة بواسطة نموذج مرشح المصدر، والذي يفترض ان الكلام يتم انتاجه عن طريق تصفية الاشارة باستخدام مرشح خطي متغير مع الزمن والشكل (6) يبين أنواع الكلام والخوارزميات المتبعة في كل نوع .



الشكل (6) انواع ترميز الكلام والخوارزميات المتبعة في كل نوع

8- الترميز التنبؤي الخطي(Linear Predication Code)LPC:

وهو يتبع لنوع النموذج الصوتي (voice coder) Vocoders حيث فيه يتكون المرمز (المشفر)من محلل ومركب. يستخرج المحلل من الكلام الأصلي مجموعة من البارامترات التي تمثل نموذج توليد الكلام، والتي يتم إرسالها بعد ذلك. ثم يعيد المركب بناء الكلام الأصلي بناءً على البارامترات المرسلة أي أخذ عينات من الكلام وتقسيمه إلى إطارات (حوالي25ms)فبدلاً من نقل الكلام الرقمي يتم بناء نموذج الكلام ثم نقل بارامترات النموذج وبعدها تجميع تقريب الكلام.

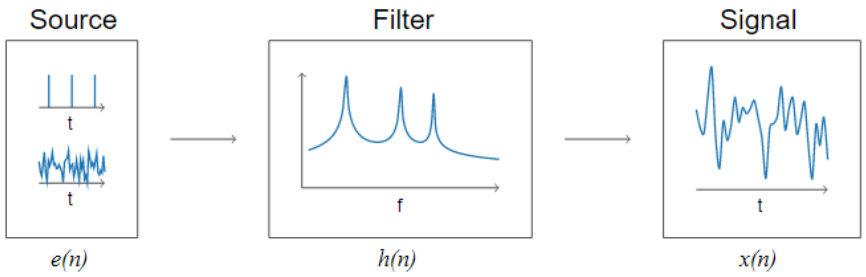
8-1-التعريف بالترميز التنبؤي الخطيLPC:

هو أداة تستخدم في معالجة الإشارة الصوتية ومعالجة الكلام، لتمثيل الغلاف الطيفي للإشارة الرقمية للكلام في شكلها المضغوط، وذلك باستخدام معلومات من نموذج تنبؤي خطي،ويعتبر من أفضل تقنيات تحليل الكلام.

يستخدم الترميز التنبؤي الخطي LPC لضغط المعلومات الطيفية من اجل التخزين والنقل الفعال وإحدى الطرق الأكثر فائدة لترميز كلام ذي نوعية جيدة في معدل بت منخفض يتضمن جزئيين رئيسيين هما: التحليل والتوليف [5] .

في قسم التحليل: يتم استخراج معامل الانعكاس من إشارة الكلام في الدخل واستخدامه لحساب المعاملات المتبقية.

في قسم التوليف: يتم إعادة بناء الإشارة الأصلية باستخدام معامل الانعكاس والمعاملات المتبقية.



الشكل (7) يوضح نموذج الترميز التنبؤي الخطي

الهدف الأصلي لـ LPC هو نمذجة إنتاج الصوت البشري والشكل (7) يوضح نموذج الترميز التنبؤي الخطي والمكون من مرشح مصدر حيث يوجد مصدر صوت يمر عبر مرشح.

يقوم المصدر e(n) بنمذجة الحبال الصوتية بينما يقوم مرشح الرنين h(n) بتشكيل القناة الصوتية والإشارة الناتجة هي: $x(n) = h(n) \cdot e(n)$ (2)

يوجد نوعان من الإشارات المحتملة للمصدر: قطار نبضي أو ضوضاء بيضاء عشوائية. هذه الإشارات نموذج لنغمات الأصوات. السمة المشتركة لكل من القطار النبضي والضوضاء البيضاء أن الغلاف الطيفي مسطحً؛ بحيث يتم نمذجة جميع المعلومات الطيفية في المرشح.

يفترض LPC أن المرشح هو مرشح متعدد الأقطاب من رتبة P^{-th} على الرغم من أنها ليست دقيقة من الناحية الفسيولوجية، إلا أنها توفر طريقة ذو مجال واسع لنمذجة الرنين. و هذا يسمح آليا بالحل عند معرفة h (n) من x(n) [5].

8-2- مشاكل الترميز التنبؤي الخطي وايجاد حلول لها :

بالنسبة لمشكلة تصميم المرشح، تماثل مشكلة تصميم الصوت باستخدام مُركِّب تناظري، يمكن التحكم في المرشحات h (n)، مع إشارة المصدر e (n)، لإنشاء نتيجة جيدة x (n).

ومع ذلك، بالنسبة للمشكلة المطروحة، فإن المعلومة الوحيدة هي x (n)، وهي الإشارة الناتجة من النظام.

وبالتالي من x(n)، نحتاج إلى استنتاج h (n) و e (n).

بالنسبة الى نظام به أقطاب [p] فقط. فإنه يحتوي على ذاكرة والعينة الحالية x (n) ستكون نتيجة الإدخال الحالي ل e (n) والعينات السابقة x (n – k)،حيث ان k = 1....P عكسيا ، هذا يعني أن h (n) تفرض علاقة بين العينات السابقة والعينة الحالية. بفرض لدينا أمثلة كافية لهذه العلاقة المفروضة، مما يعطينا الامكانية على استخراج بعض المعلومات حول المرشح.

يمكننا صياغة العلاقة بين المدخلات والمخرجات على النحو التالي،

$$(3)\ X(Z)=H(Z)E(Z)=\frac{1}{1-\sum_{k=1}^{p}a_k z^{-k}}E(Z)$$

ويمكن التعبير عن(z) X من خلال المدخلات (z) E والعينات السابقة وفق الآتي:

$$X(Z)=\sum_{k=1}^{p}a_k z^{-k}X(Z)+E(Z) \qquad (4)$$

$$x(n)=\sum_{k=1}^{p}a_k x(n-k)+e(n)\ (5)$$

الآن علينا إيجاد المعامِلات a_k ، $k=1....p$. من جديد ،و نحاول ايجاد تردد الرنين في حال كان لدينا عينات N>>p، فعندئذٍ نحصل على N معادلة لـ a_k ، وبالتالي فإن a_k محدد من خلال تحويل المعادلة (5) إلى شكل مصفوفة أحادية البعد وفق الآتي:

$$x(n-p+i)\in 1\times p\ \ldots \text{For}\ x_i=\left[x(n-1+i).....x(n-p+1\right]\in R_1\times p_{x_i}=x(n-1+i)$$

$$a=\begin{bmatrix}a_1\\ .\\ .\\ .\\ a_p\end{bmatrix}\in R_P\times 1$$

ومنه نحصل على N من خلال مجموعة المعادلات التالية:

$$x(n)=x_0a+e(n)$$
$$x(n+1)=x_1a+e(n+1)$$
$$.$$
$$. \qquad (6)$$
$$.$$
$$x(n+N)=x_Na+e(n+N)$$

نحدد قيمة المتغيرات وفق الآتي :

$$\mathbf{b}=\begin{bmatrix}x(n)\\ \vdots\\ x(n+N)\end{bmatrix}\in\mathbb{R}_{N\times 1},\ \mathbf{e}=\begin{bmatrix}e(n)\\ \vdots\\ e(n+N)\end{bmatrix}\in\mathbb{R}_{N\times 1}\ \text{and}\ \mathbf{A}=\begin{bmatrix}\mathbf{x}_0\\ \vdots\\ \mathbf{x}_N\end{bmatrix}\in\mathbb{R}_{N\times p},$$

ثم نقوم باستبدال المتغيرات في المعادلة (6) وإعادة ترتيب حدود ضرب المصفوفة، ونحصل عليها كالاتي :

$$e=b-Aa.....(7)$$
$$=b-\hat{b}.....(8)$$
$$\equiv A\cdot a\hat{b}$$

تُعرف المعادلة أعلاه بالانحدار الخطي، حيث $\hat{b}$تقدير لقيمة b بالاعتمادً على قيمةA مع الأوزان a وe تقدير الخطأ المتبقي.

وكما ذكرنا من قبل، فإن المعادلة (6) محددة بشكل مبالغ فيه ولا يوجد حل دقيق، يمكننا فقط محاولة إيجاد "توافق جيد". الحل الملائم لهذه المشكلة هو المربعات الصغرى، والحل المفضل لـ LPC من اجل حل المربعات الصغرى هو اختيار قيم a_k التي تقلل من ||e(n)||2، قوة e المتبقية وفق الحل الآتي :

$$a = A^t b$$
$$= (A^T A)^{-1} A^T b.....(9)$$

بعد إيجاد قيم(a)ومن خلال المعادلة (6) ، يمكننا حساب القيم المتبقية(e)من خلال نمذجة الرنين لا بد من وجود قطار من النبضات بتردد معين، أو إشارة ضجيج عشوائية أو مزيج من الاثنين. في حال لم يتحقق ذلك، فسوف يتداخل بعض الرنين مع e، ولن يكون مسطحًا طيفيًا.

في البداية لا بد من تحديد فيما إذا كانت الإشارة الصوتية (رنين) أو إشارة ضجيجية وذلك من اجل تحديد درجة الصوت. هناك طرق كشف بسيطة باستخدام الطرق الصفرية وطرق الارتباط التلقائي. إما الطرق الأكثر تعقيدًا فتستخدم أجهزة التصفية الذكية أو النماذج الإدراكية (cepstrums). بالنسبة للحالة المنحدرة نحتاج إلى قوة إشارة المصدر مع التردد الأساسي. ما بالنسبة للإشارة الضجيجية نحتاج إلى قياس تباين العينات.

من المعروف أن ثلاث ترددات للإشارة الصوتية تمكننا من تمييز معظم الحروف. وبالتالي فإن ضبط p = 6 سيعمل بشكل جيد مع إشارة الكلام. وهو أحد الأسباب التي تجعل LPC يعمل بشكل جيد مع الكلام. يعد اختيار القيمة المثلى p لأنواع إشارات أخرى من الإشارات مشكلة غير محددة [6].

نقوم بعملية التحليل وإعادة تركيب الإشارة حيث وجدنا المعاملات a_k ولدينا نموذج للإشارة الصوتية، ومع ذلك فإن الحل أثناء إيجاد تردد الرنين سينتج صوتًا رنانًا واحدًا فقط، مثل "آه"، وليس الجملة الكاملة "أنا رائع هكذا".

تختلف إشارات العالم الحقيقي عن بعضها البعض، مثل الموسيقى والكلام، بمرور الوقت. لذلك نحتاج إلى تقطيع الإشارة إلى أجزاء صغيرة و نمذجة كل جزء باستخدام المعادلات المستخدمة في إيجاد تردد الرنين.

تستخدم الطريقة المعتمدة في معالجة الصوت تقنية تراكب إضافة OLA، بحيث نمثل الإشارة وفق المعادلة الآتية :

$$\sum_{n=-\infty}^{\infty} w(n - m \times R) = 1 \qquad (10)$$

وهي تمثل معادلة OLA الثابتة، حيث يُطلق على R حجم الخطوة. وبالتالي بدلاً من تشغيل LPC على الإشارة الكاملة x (n)، نستخدم نسخة إطارات $x_m(n) = w(n) \times x(n + mR)$ ونظرًا لخاصية التداخل والإضافة الثابتة ، فإن إضافة الإشارات ذات الإطارات ستعيد الإشارة الأصلية وفق المعادلة الآتية [6].

$$\sum_{n=-\infty}^{\infty} w(n - m \times R) x(n) = x(n) \qquad (11)$$

أي قيمة خارج النافذة بطول N تساوي الصفرً. هذا يعني بالنسبة إلى x_i في المعادلة (6) :

$x(n-m+i)=0$ إذا كان $n-m+i \prec 0$ أو $n-m+i \le N$ حيث $m=1.....p$

نذكر هذا لأن المصطلحات في المعادلة (8) أصبحت ارتباطًا تلقائيًا وتُعرف أيضًا باسم معادلات Yule–Walker. تسمح هذه الطريقة باستخدام FFT لحساب الارتباطات التلقائية.

مثال يوضح استخدام LPC لإشارة الكلام. باختيار نافذة بطول240 = N عينة، وهو طول نافذة 30ms عند8KHz، وهي قيم مناسبة لإشارة الكلام. وباستخدام 50٪ OLA، مما يعني أن عينة120 = 0.5N = R وبما أن الإشارة كلام فقط نهتم بـ ثلاث قمم (صيغ)، لذلك نختار 6 = p لكل خطوة 120 عينة، بعد تطبيقLPC، يتبقى لنا 7 عينات ،المعامِلات الستة a_k والمتغير المتبقي δ^2 والنتيجة أننا خفضنا كمية المعلومات من 120 عينة إلى 7 (حوالي 17 إلى 1). أي أننا رمزنا إشارة الصوت الأساسي بشكل مضغوط.

لفك تشفير الإشارة ، نقوم بتشغيل المعاملات من خلال نموذج LPC كما في الشكل (7)وذلك باستخدام التباين δ2 للتحكم في المصدر، والمعاملات a_k للمرشح، نحصل على إشارة تقديرية وفق المعادلة الآتية :

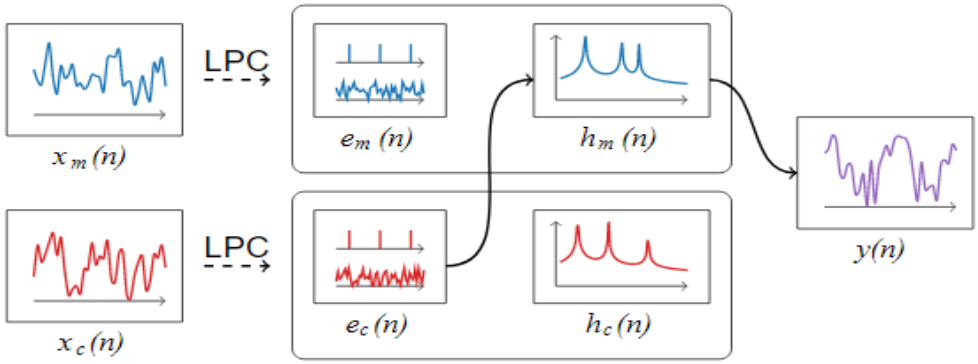
$$m(n)=w(n)(e(n)\cdot h(n))\hat{x} \qquad (12)$$

للحصول على الإشارة الكاملة $\hat{x}(n)$تقوم بالتداخل مع إضافة إشارات النوافذ التي تم تشفيرها او ترميزها.

بعد ذلك نقوم بعملية فك الترميز حيث ان النظام المتبع هو نظام خطي لذلك نستخدم طريقة المربعات الصغرى التي تحل المعادلات الخطية ونستخدم القيم السابقة ل $x(n)$خطيًا لإيجاد المعاملات a_k التي تمثل أفضل توقع للقيمة الحالية للترميز.

تم شرح كيفية استخدام LPC لتحليل وإعادة تركيب إشارات الكلام من خلال عملية الترميز وفك الترميز.

وبالتالي يمكن اعتبار LPC خوارزمية ترميز، من خلاله يمكننا تقديم مقارنة مهمة بين ترميز الصوت وفق خوارزميات MP3 و الترميز باستخدام LPC الذي يقوم بنمذجة المصدر بينماMP3 تقوم بنمذجة الوجهة.



الشكل (8) يوضح عملية التوليف التبادلي في الترميز التنبؤي الخطي

بالنظر إلى LPC من منظور التحليل / التوليف بتطبيقات مختلفة. وخير مثال على ذلك هو التوليف التبادلي كما في الشكل (12) بعد تشغيل إشارة من خلال LPC ، يتبقى لنا نموذج المرشح ، h (n) ونموذج المصدر e (n)

يتحكم المرشح h(n) في صدى أو نغمة الإشارة ، بينما يتحكم المصدر e(n) في النطق. إذا قمنا بتشغيل إشارتين مختلفتين من خلال LPC ، فسيكون من الممكن استخدام مصدر نموذج واحد (الحامل $e_c(n)$) ومرشح النموذج الآخر (المعدل $h_m(n)$)لإنشاء نموذج مختلف ،

وتكون اشارة الخرج $y(n) = h_m(n) \cdot e_c(n)$ وبالتالي يكون لدينا توليف متبادل حيث نحصل على نتيجة جيدة وذلك عند أخذ المرشح إشارات الكلام والمصدر ذو صوت واضح الطيف .

8-3- القيود التي يفرضها استخدام الترميز التنبؤي الخطي :

مع كل مزايا LPCهناك قيود حيث يجب أن تتناسب الإشارة الصوتية بشكل مقبول مع نموذج مرشح المصدر،و إذا كانت الاشارة المتشكلة من عدة اصوات تكون معالجتها غير جيدة.

يمكن تحقيق مزيد من التحسينات في جودة الصوت دون زيادة كبيرة في البيانات المطلوبة باستخدام أكواد الإثارة. يُطلق على هذا النهج اسم التنبؤ الخطي المثار بالشفرة (الترميز) CELP،

إن LPC نفسه غير مناسب للأصوات ذات العديد من الإشارات العابرة وذلك في حال أن الإشارة ثابتة داخل مقطع الإطار مما يؤدي إلى تشوه بعرض نوافذ الهجوم العابر .

لمعالجة هذه المشكلة تم إيجاد طريقة التنبؤ الخطي للتردد الزمني المتتالي CTFLP ، حيث تتم نمذجة الغلاف الزمني لإشارة المصدر بعد عملية LPC في المجال الزمني.

مع كل هذه القيود يعتبر LPC طريقة جيدة لتحليل ونمذجة الأصوات [8].

9-دراسة الترميز التنبؤي الخطي من خلال Simulink الماتلاب :

نفرض ان لدينا إشارة كلام بتردد 8KHz بحيث تمثل إشارة الدخل والتي سنقوم بتمثيلها في برنامج الماتلاب من خلال بلوك (from multimedia file) والذي يتيح لنا التعبير عن إشارة الدخل المذكورة.

والترميز التنبؤي الخطي يتكون من جزأين كما أسلفنا أولهما التحليلي والمبين في الشكل (9) والثاني التوليفي كما في الشكل (12) وأيضا عملية التكميم الموضحة في الشكل (10).

9-1-القسم التحليلي :

نحتاج إلى مرشح تعزيز(Pre-Emphasis Filter) والذي يعمل على تعزيز المركبات الترددية العالية من إشارة الكلام والتي يتم فقدها أثناء إنتاج الكلام ويقلل المرشح من طيف التردد للإشارة التي تمر بها حيث يتم تمثيل

Pre-Emphasis Filter في الماتلاب من خلال بلوك (Discrete FIR Filter).

نحتاج لتطبيق عملية التأطير(FRAMING)حيث يتم تقسيم إشارة الكلام بعد مرورها بمرشح التعزيز إلى إطارات وكل إطار يحتوي N عينة من الكلام ومن ثم فصل الإطارات المجاورة لبعضها البعض وتسمى هذه

الطريقة بتأطير الإشارة ويمكن تمثيل هذه العملية من خلال بلوك (Buffer) والتعديل بخصائصه ليصبح مناسب للعملية.

نحتاج لتطبيق عملية نافذة هامينغ WINDOWING لتقليل انقطاع الإشارة في بداية ونهاية كل إشارة ونستخدم من اجل هذه العملية نافذة هامينغ على كل إطار على حدى بالمعادلة التالية (13) التي تحدد عدد العينات في N إطار.

$$w(n) = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}).....,0 \prec n \prec N-1 \qquad (13)$$

حيث nعدد العينات في الإطار

ويمكن تمثيل هذه العملية من خلال بلوك (window function) والتعديل بخصائصه من خلال اختيار نوع النافذة هامينغ.

نحتاج لتطبيق عملية الارتباط التلقائي(Autocorrelation)من خلال المرمز الذي يعمل على تقليل متوسط الخطأ التربيعي من أجل إيجاد معامل المرشح الذي يتطابق بشكل أفضل مع الجزء الذي يتم تحليله ويتم تمثيل هذه العملية من خلال بلوك (Autocorrelation) والتعديل بخصائصه ليصبح مناسب للعملية.

نحتاج لتطبيق عملية (Levinson Durbin Algorithm) وهي خوارزمية تستخدم لتحديد معامل LPC

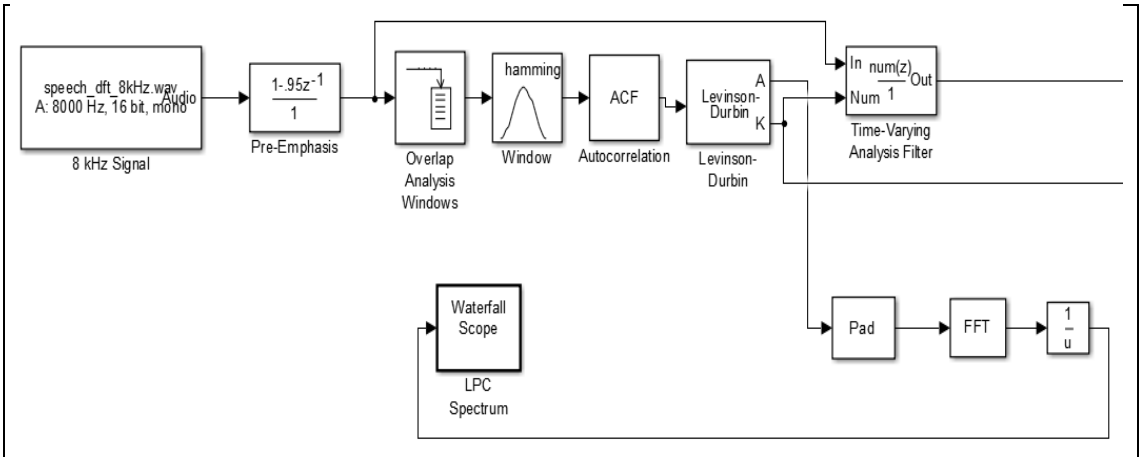
ويمكن تمثيل هذه العملية من خلال بلوك(Levinson–Durbin) والتعديل بخصائصه ليصبح مناسب للعملية.

نحتاج لتطبيق (Time Varying Analysis Filter) وان هذا النوع من المرشحات هو مرشح صفري يتم من خلاله تمرير إشارة الكلام التي هي على خرج المرشح (Pre–Emphasis Filter) لإنشاء الإشارة المتبقية

(إشارة التنبيه) ويكون عرض النطاق الترددي للإشارة المتبقية أقل من الإشارة الأصلية.ويمكن تمثيل هذه العملية من خلال بلوك (Discrete FIR Filter) والتعديل بخصائصه.

ثم يتم إظهار الطيف الترددي لإشارة الكلام وذلك في خرج الجزء التحليلي من الترميز التنبؤي.

الإشارة على خرج Levinson Durbinهي إشارة في مجال الزمن ونحتاج إلى تحويلها للمجال الترددي وذلك من اجل إظهار الطيف الترددي للإشارة الناتجة بعد الترميز ويتم عرض الطيف الترددي للإشارة من خلال بلوك (waterfall scope) والتعديل على بعض خصائصه[9].

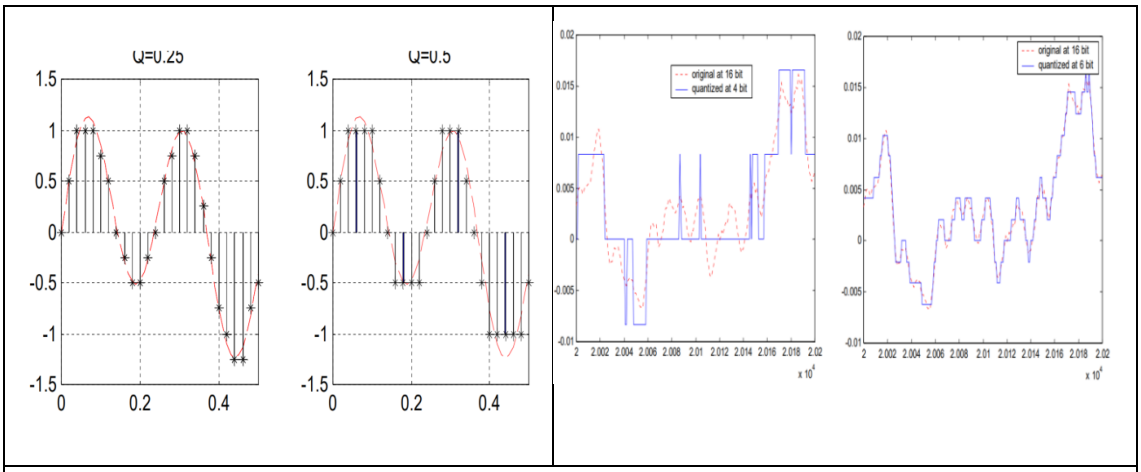


الشكل (9) يوضح القسم التحليلي من دارة الترميز التنبؤي الخطي

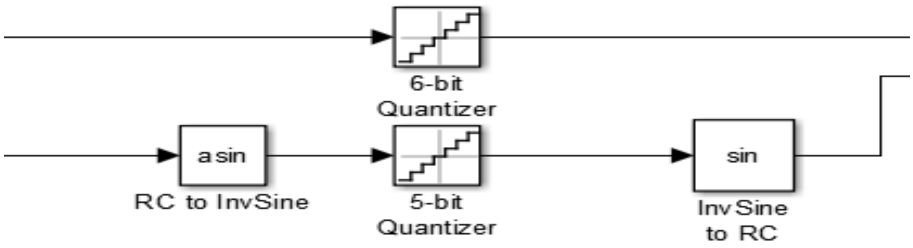
9-2- قسم التكميم :

بعد الجزء التحليلي يجب ان نقوم بعملية التكميم(Quantization). حيث يشار إلى الفرق بين قيمة المدخلات والقيم بعد التكميم بخطأ التكميم.

والشكل (10) يبين الطريقة التي يتم من خلالها عملية التكميم ومقسم التكميم في دارة الترميز التنبؤي موضح بالشكل (11) .



الشكل (10) يوضح الطريقة التي يتم خلالها التكميم



الشكل (11) يوضح قسم التكميم من دارة الترميز التنبؤي الخطي

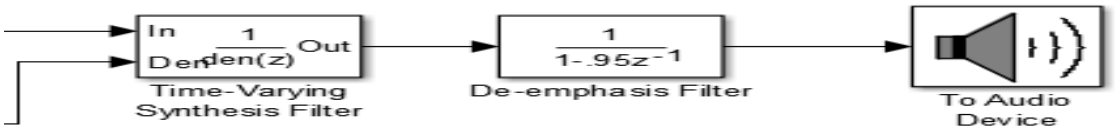
9-3- قسم التوليف :

وهو عكس الجزء التحليلي حيث ان كل بلوك يعبر عن عملية فك الترميز(الشفرة) بشكل فردي وتسلسل مقاطع الصوت المستنسخة يتم ضمها معًا لتمثيل إشارة الكلام المدخلة بالكامل.

نستخدم (Time Varying Synthesis Filter) لإعادة تشكيل إشارة الدخل الأصلية(الأولية) ذات المركبات الترددية العالية المعززة بناءً على الإشارة المتبقية ومعامل الانعكاس.

نستخدم (De-Emphasis Filter) مرشح إزالة التعزيز وهو مرشح رقمي يستخدم لاستعادة مكونات التردد العالي إلى ما قبل التعزيز في الإرسال حيث إن الطريقة تتم من خلال بلوك(filter allpole filter) من

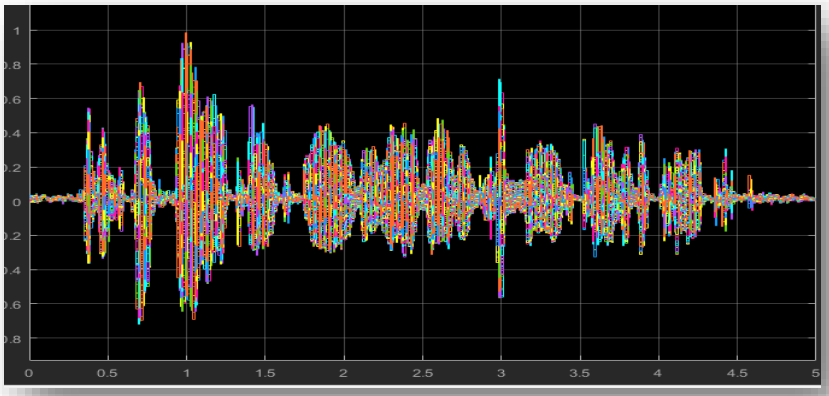
خلال التعديل على خصائصه ليصبح مناسبا من خلال اختيار بنية المرشح الداخلي في البلوك (Direct From) [8] والشكل (12) يوضح القسم التوليفي في الترميز التنبؤي الخطي لإشارة الكلام.



الشكل (12) يوضح قسم التوليف من دارة الترميز التنبؤي الخطي

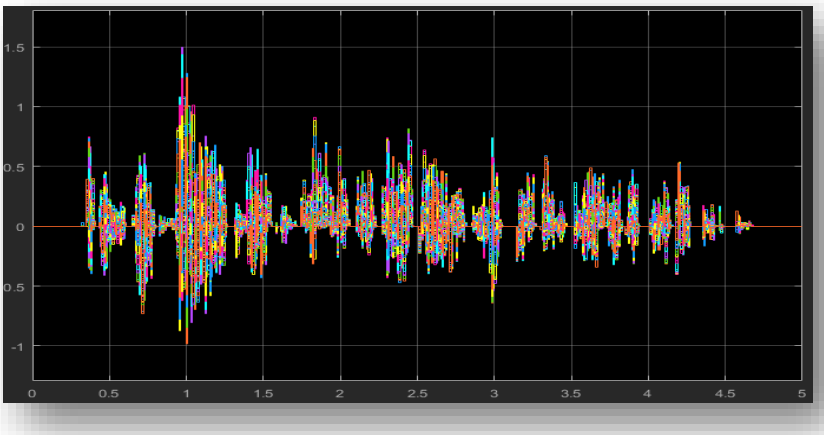
9-4- النتائج :

بعد ان قمنا بدراسة مكونات الترميز التنبؤي الخطي وبعد تنفيذ عملية المحاكاة باستخدام برنامج الماتلاب كانت النتائج التي حصلنا عليها كالتالي:

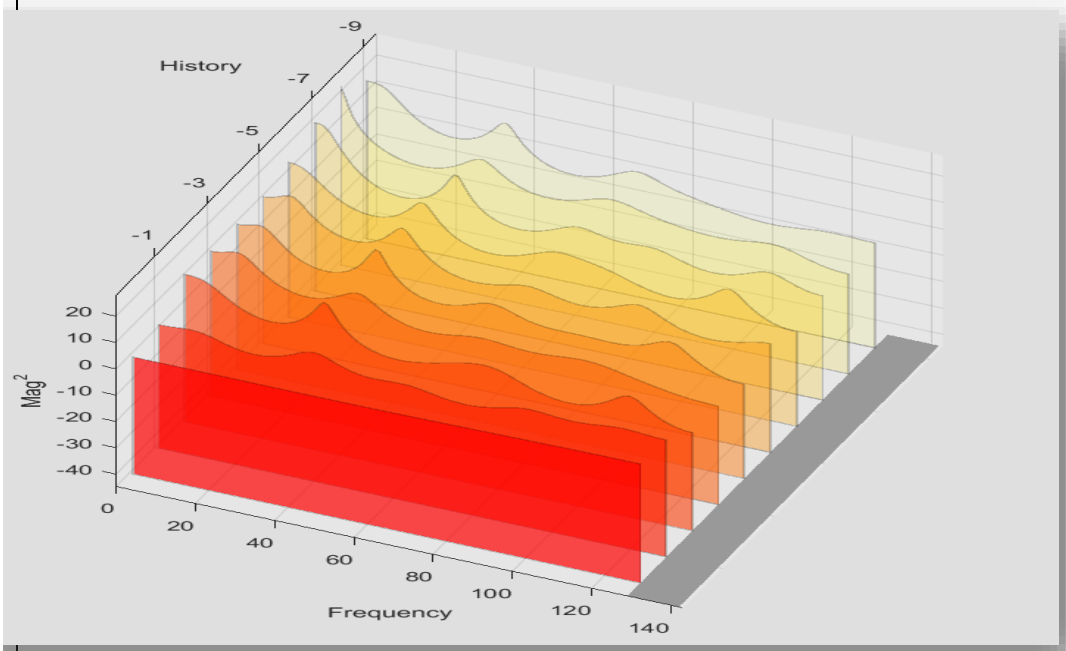


الشكل (13) يوضح الاشارة الصوتية في الدخل

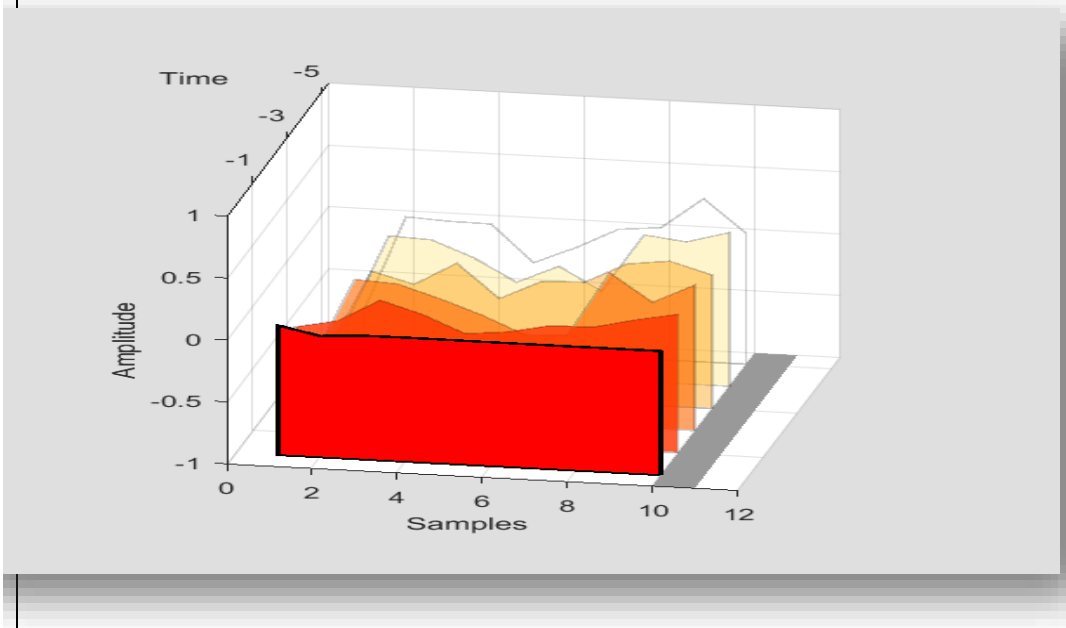
نلاحظ من الشكل (13) ان اشارة الدخل تحوي ضجيج وتشويش يقابله في الشكل (14) اشارة الخرج خالية تقريبا من أي ضجيج وتشويش .



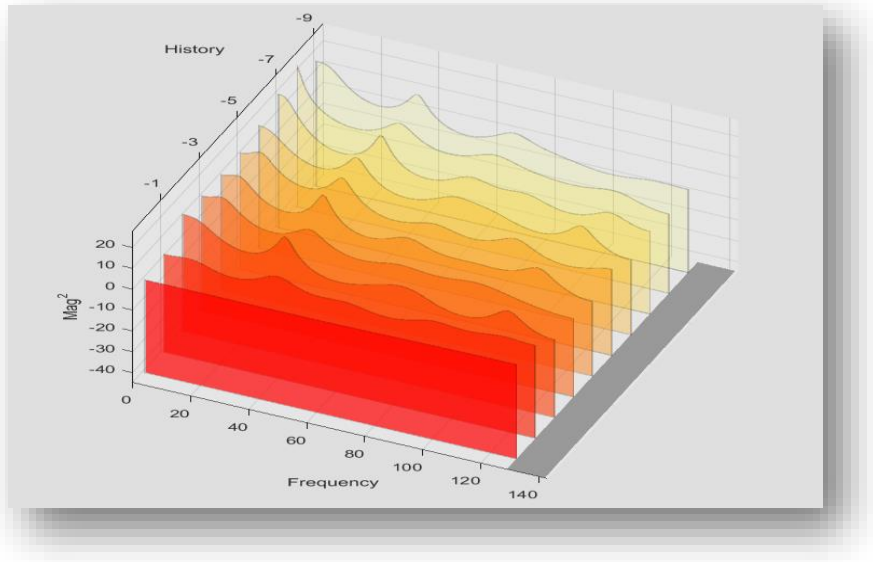
الشكل (14) يوضح الاشارة الصوتية في الخرج



الشكل (15) يوضح المخطط الطيفي للترميز التنبؤي الخطي



الشكل (16) يوضح معاملات الانعكاس



الشكل (17) يوضح الاشارة المتبقية

يوضح الشكل (15) المخطط الطيفي للترميز التنبؤي الخطي الذي يبين لنا أهميته من خلال ضغط الصوت وذلك من خلال إرساله عبر قناة ضيقة .

تم تحليل إشارة الكلام عن طريق تقدير معاملات الانعكاس المتمثلة بعينات samples و مطال إشارة الكلام amplitude كما هو موضح في الشكل (16).

تسمى عملية إزالة الصيغ بالترشيح العكسي ، وتسمى الإشارة الناتجة عن طرح الإشارة الأصلية المنمذجة من معاملات الانعكاس بالإشارة المتبقية كما هو موضح في الشكل (17) .

9-10-الاستنتاجات والتوصيات:

1. تم تحليل إشارة الكلام عن طريق تقدير معادلات معاملات الانعكاس.
2. إيجاد الحلول لمشاكل LPC، وإزالة آثارها من إشارة الكلام.
3. توصلنا إلى إشارة خرج ذو جودة أعلى وحجم اقل ووضوح صوتي اكبر وإبراز الترددات المهمة في الإشارة الأصلية.
4. الترميز التنبؤي الخطي هو طريقة فعالة لضغط الصوت وذلك من خلال إرساله عبر قناة صوتية ضيقة وهذا أفضل اقتصاديا وعمليا.
5. نلاحظ أن الترميز التنبؤي الخطي ذو فائدة جيدة لترميز الكلام في معدل بت منخفض، ويقدم تقديرات عالية الدقة لإشارة الكلام وذلك لأنه يعتمد على تمثيل الغلاف الطيفي لإشارة الكلام الرقمية في شكلها المضغوط.
6.

10- المراجع :

[1]O'Shaughnessy, Douglas. "Linear predictive coding." IEEE potentials 7.1 (1988): 29-32.

[2] Bradbury, Jeremy. "Linear predictive coding." Mc G. Hill (2000).

[3] Wijoyo, Suryo, and S. Wijoyo. "Speech recognition using linear predictive coding and artificial neural network for controlling movement of mobile robot." proceedings of 2011 international conference on information and electronics engineering (ICIEE 2011).

[4] Javier, Rios Jesus, and Youngwook Kim. "Application of linear predictive coding for human activity classification based on micro-Doppler signatures." IEEE Geoscience and Remote Sensing Letters 11.10 (2014): 1831-1834.

[5] Liu, Peng, Songbin Li, and Haiqiang Wang. "Steganography integrated into linear predictive coding for low bit-rate speech codec." Multimedia Tools and Applications 76.2 (2017): 2837-2859.

[6Fadhilah, Davita Nadia, Rita Magdalena, and Sofia Sa'idah. "INDIVIDUAL IDENTIFICATION SYSTEM DESIGN THROUGH VOICE USING LINEAR PREDICTIVE CODING METHOD AND K-NEAREST NEIGHBOR." JurnalTeknikInformatika (Jutif) 2.2 (2021): 95-100.

[7] Ofner, André, Johannes Schleiss, and Sebastian Stober. "Hierarchical Predictive Coding and Interpretable Audio Analysis-Synthesis."

[8] Terenzi, Terenzi. "Adaptive Motor Control: a Predictive Coding Approach." (2021).

[9] Liu, Peng, Songbin Li, and Haiqiang Wang. "Steganography integrated into linear predictive coding for low bit-rate speech codec." *Multimedia Tools and Applications* 76.2 (2017): 2837-2859.