

كراسات المركز

سلسلة يصدرها مركز البحث العلمي

والتقني لتطوير اللغة العربيّة

مفاهيم حول تفسير الكلام

استعمال التّحليل إلى القيم الشّاذّة

مراد عبّاس

العدد السادس

2007

ملخص

تعتبر ترجمة هذه الدراسة محاولة متواضعة للإسهام في إثراء الرصيد العلمي العربي في مجال تشفير الكلام. وقد حاولنا بقدر الإمكان تقريب المادة العلمية من خلال الاشتغال على اللغة العربية وانتقاء من المصطلحات بما يتوافق مع الأسيقة المختلفة لهذه الدراسة.

وقد قمنا في هذه الدراسة بمعالجة تشفير الكلام لما يكتسيه من أهمية في مختلف الميادين التطبيقية.

ولتحقيق الأهداف المسطرة في هذه الدراسة قدمنا تعريفا للمفاهيم الأساسية المستخلصة من نظرية المعلومات وكذا المراحل الأولى التي تسبق التشفير؛ كالتجزئة إلى عينات والتقسيم الكمي. وجنحنا بعد ذلك إلى عرض بيان حثيات تجربتنا وتحليل نتائجها والتعليق على معطياتها المتعلقة باستعمال المحولة SVD في ضغط الكلام.

إن الطريقة المعتمدة مبنية على دمج صنفين من التشفير : المحولة SVD والتشفير الوسيط، الذي يرتبط بمعاملات التنبؤ الخطي التي تشهد استعمالا واسعا في ميدان العلاج الآلي للكلام.

ومن الواضح أن أنظمة التشفير التي تقوم أساسا على استعمال التحويل تمثل العينات بمجموعة من المعاملات والتي بدورها تخضع لعملية التقسيم الكمي، ثم يتم بعد ذلك إرسالها.

أدرجنا في هذه التجربة معاملات التنبؤ الخطي في نظام التشفير بحيث يتم تطبيق المحولة على هذه المعاملات بدلا من العينات. ويعد استعمال

التقسيم الكمي الشعاعي إجراء ضروريًا، نظرًا لكثرة الأشعة المحتوية على المعاملات، وقد وقع اختيارنا على طريقة LBG.

وبناء على استخدام المعاملات المقسمة كميًا، قمنا باستعادة الإشارة المشفرة عن طريق مرشح إصدارٍ محرّضٍ بالإشارة المتبقية.

الكلمات المفتاحية : تشفير الكلام، المحولة SVD، التقسيم الكمي الشعاعي.

الفهرس

9	1. مقدمة
10	2. مفاهيم حول نظرية التشوه
10	1.2. درجة التعادل لمنبع
12	2.2. نظرية التشوه
16	3. التمثيل الرقمي (الرقمنة) للإشارات الصوتية
17	1.3. تعريف مبسط للرقمنة
20	4. قياس نوعية المشفر
20	1.4. القياس الموضوعي
20	2.4. القياس الذاتي
21	5. مفهوم التحويل
24	1.5. تفسير
25	6. التشفير باستعمال التحويل
27	1.6. الإرفاق الديناميكي للأرقام الثنائية
28	7. التحويل SVD
29	1.7. القيم والأشعة الصرفة
30	2.7. التحليل إلى قيم شاذة لمصفوفة
31	3.7. تفسير التحويل SVD
32	8. التشفير باستعمال SVD
34	9. التقسيم الكمي الشعاعي
34	1.9. تعريف
35	2.9. التقسيم الكمي الشعاعي الإحصائي
36	3.9. التقسيم الكمي الشعاعي الجبري
37	4.9. خوارزمية LBG
38	5.9. كفاءة المقسم الكمي الشعاعي
44	10. تجارب ونتائج
50	11. الخاتمة
53	ثبت المصطلحات (عربي / فرنسي)

1. مقدمة

لقد كان الكلام ولا يزال أهم وسيلة للتخاطب بين بني البشر. وتسعى البحوث حاليا لتمكين الإنسان من التخاطب مع الآلة. وهذا ما يستوجب التحكم في مجالات جوهرية ثلاث وهي : التركيب الاصطناعي للكلام، التعرف الآلي على الكلام والتشفير.

ليس تشفير إشارة الكلام سوى وضعها في شكل معين، ونعني بذلك تمثيلها تمثيلا خاصا بهدف التخزين أو الإرسال. وتقتضي الأصوات مساحة هائلة لاستيعاب التخزين وتدفقات إرسال عالية، فلأجل ذلك غدا ضغطها ضروريا، تكون الغاية منه شغل حيز صغير والوصول إلى تدفق أدنى. ولهذه الاعتبارات، يتم اللجوء إلى التقليل من كمية المعلومات الموجودة في إشارة الكلام وذلك باستغلال التكرار الكلامي الذي يميزها، أو بتكليف المحتوى المعلوماتي للقدرات المحدودة للمستعمل (الأذن البشرية).

من هذه المنطلقات، عرضنا في الفصل الثاني مفهومي نظرية المعلومات ونظرية التشوه. كما أحطنا بالمرحل الأساسية التي تجتازها إشارة الكلام قبل التوقف عند محطة التشفير. ونظرا لأهمية التحويل والتعامد، قمنا بعرض مثال يصف هاتين القاعدتين متوخين في ذلك التبسيط.

قدمنا في هذه الدراسة أيضا تعريفا للمجولة SVD يشمل الخصائص التي تميزها والتي تجعلها أداة فعالة في التشفير. وقد ألحقنا هذا التعريف بشرح التشفير الذي ينهض على مبدأ استعمال التحويل بصفة عامة، ثم

باستعمال التحويل SVD، حيث أدرجنا مدى كفاءة المقسم الكمي الشعاعي بوضع منبع "غوسي" عند مدخله.

في المرحلة الأخيرة من هذه الدراسة عرضنا نتائج التجارب المطبقة على نظام التشفير هذا وقمنا بتقييمها باستخدام قياسات موضوعية وأخرى ذاتية.

2. مفاهيم حول نظرية التشوّه

1.2. درجة التعادل لمنبع

تعرف درجة التعادل لمنبع $X=(x(0), x(1), \dots, x(N-1))$ ذي زمن متقطع، مستقر، ودون ذاكرة، بأنها المعلومة-الذاتية المتوسطة لهذا المنبع [1] وتعطى بالعلاقة التالية :

$$H(X) = E[I(X)] \quad (2.1)$$

حيث تعتبر $I(X)$ المعلومة-الذاتية للمنبع، ويمكن الحصول عليها بالمعادلة (2.2) :

$$I(X) = - \log_2 P(X) \quad (2.2)$$

تأخذ كل عينة قيمها من المجموعة $\{a_0, a_1, \dots, a_{N-1}\}$ ويرفق لها الاحتمال $p(x(i) = a(i) = p_X(a_i)$.

إن احتمالات مختلف عناصر المنبع X معرفة كما يلي :

$$p(a_0), p(a_1), \dots, p(a_{N-1})$$

وبالتالي فالاحتمال $P(X)$ يسمح بتعريف المعلومة الناتية. وانطلاقا

من (2.1) و(2.2) تصبح عبارة درجة التعادل كالتالي :

$$H(X) = -E[\log_2 P(X)] = -\sum_{i=0}^{N-1} P_X(a_i) \log_2 P_X(a_i) \quad (2.3)$$

علما بأن :

$$0 \leq P_X(a_i) \leq 1 \quad \sum_{i=0}^{N-1} P_X(a_i) = 1 \quad (2.4)$$

إن درجة التعادل لمنبع هي دائما موجبة ومحدودة:

$$0 < H(X) < \log_2 N$$

عندما تكون $H(x) = \log_2 N$ نحصل على احتمال منتظم $P_X = 1/N$

ويكون المنبع عندئذ غير قابل للتنبؤ كليًا. ويكون هذا الأخير عكس ذلك إذا كانت $H(x)$ معدومة.

أما في حالة وجود متغيرين عشوائيين X و Y مرفقين بالقيم على

الترتيب : $\{a_0, a_1, \dots, a_{N-1}\}$ و $\{b_0, b_1, \dots, b_{N-1}\}$ مع توزيع للاحتمال

المرفق $p_{XY}(a_i, b_j)$ فإن درجة التعادل الشرطية المستخلصة من نظرية

الاحتمال الشرطي تتيح معرفة المعلومة الخاصة بمنبع ما، شريطة

معرفة المنبع الآخر. ويمكن حسابها باستعمال المعادلة (2.5) :

$$H(X, Y) = -E[\log_2 P_{Y/X}(X/Y) = -\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_{X/Y}(a_n, b_m) \log_2 P_{Y/X}(a_n / b_m) \quad (2.5)$$

2.2. نظرية التشوه

باستعمال المقسم الكمي يحدث تقريب إشارة المدخل X بإشارة أخرى Y وهو ما يسمى بالتشوه. تتيح العلاقة التالية معرفة معيار الالتزام، وهو مقدار التشوه الحاصل.

$$\rho_k(x, y) = \frac{1}{k} \sum_{i=0}^{k-1} \rho_k(x(i), y(i)) \quad (2.6)$$

تعتبر الكمية ρ إذا قياسا للتشوه الناتج بين مركبات الشعاعين X و Y ، ويمكن التعبير رياضيا عن التشوه المتوسط المتعلق بدالة الاحتمال المرفق للانتقال بالعبارة الآتية :

$$d(P_{Y/X}) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_X(a_n) P_{Y/X}(b_m / a_n) \rho(a_n, b_m) \quad (2.7)$$

وللعلم، فإن العلومة المتبادلة هي كميّة المعلومات للإشارة Y والتي تعتمد أساسا على ملاحظة قيم الإشارة X . ولحسابها تستعمل المعادلة (2.8) :

$$I(P_{\gamma/x}) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_x P_{\gamma/x} (b_n/a_n) \log_2 \frac{P_{\gamma/x} (b_n/a_n)}{P_x (b_n)} \quad (2.8)$$

$$P_{\gamma/x} = \sum_{n=0}^{N-1} P_x(a_n) P_{\gamma/x} (b_n/a_n) \quad (2.9)$$

إن دالة التشوه $R(D)$ الممثلة بالعلاقة (2.10) تسمح بالحصول على التدفق الأدنى R مع التناهي عن وجود التشوه D . وعلى العكس من ذلك نستطيع تحديد قيمة معينة للتدفق بغية الحصول على أدنى قيمة للتشوه.

$$R(D) = \min_{P_{\gamma/x} \in P_D} I(P_{\gamma/x}) \quad (2.10)$$

$$\text{avec } P_D = \left\{ P_{\gamma/x} : d(P_{\gamma/x}) \leq D \right\}$$

إن الشفرة C ذات حجم N وطول قطعة K هي زمرة من N مجموعة. وكل مجموعة ذات بعد K .

هذه المجموعات هي مجموعات ثنائية تسمى كلمات ذات شفرة، ويقدر طولها بالقيمة $\log_2 N$. فمثلا إذا أردنا تشفير شعاع مكون من K عينة، فيكون التدفق R للعينة الواحدة على النحو التالي :

$$R = \frac{\log_2 N}{k} \quad (2.11)$$

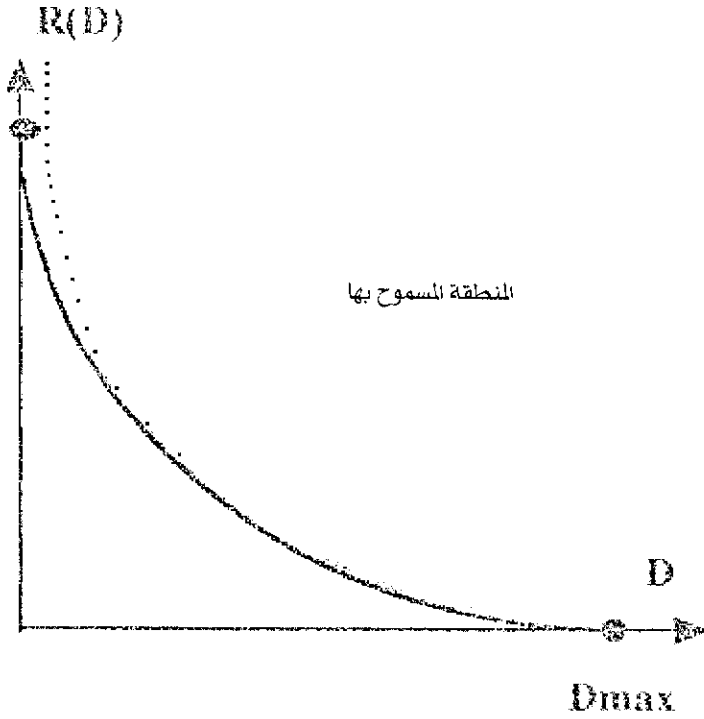
تعتبر دالة التشوه $R(D)$ دالة مستمرة، رتيبة ومتناقصة. فهي تمثل الحد النظري لكفاءة أنظمة التشفير. ويمكن تمثيلها بالعلاقة (2.12) :

$$R(D) = \lim_{k \rightarrow \infty} \frac{1}{k} \log_2 N \quad (2.12)$$

في الواقع، يمكن الحصول على كفاءة قريبة من هذا الحد. ولتوضيح الرؤية أكثر فيما يتعلق بدالة التشوه والحد النظري الذي تمثله ارتأينا أن نرسمها في الشكل (1.2).

إن الرسم المبين في الشكل (1.2) يوضح جلياً أنه توجد منطقة مسموح بها، محددة بالنهاية الصغرى للدالة $R(D)$ تسمى بحد "شانون" المحظور تخطيه. حيث يمكن ملاحظة أن $R=H$ في حالة انعدام التشوه. أما إذا كانت قيمة هذا الأخير قصوى فنحصل على $R=0$.

في حالة منبع مستمر يؤول المنحنى $R(D)$ إلى ما لا نهاية إذا كان التشوه ضعيفاً.



الشكل (1.2) . دالة التشوه $R(D)$ في حالة الزمن المتقطع (بالخط المستمر) وفي حالة منبع مستمر (بالخط المنفصل).

3. التمثيل الرقمي (الرقمنة) للإشارات الصوتية

لكي يتم تمثيل إشارة رقمياً يجب إخضاعها أولاً لعملية التجزئة إلى عينات، مع اختيار ملائم لتواتر التجزئة المستخدم F_s يراعي ما يسمى بشرط "شانون". ثم يطبق التقسيم الكمي بعد ذلك على كل عينة، حسب عدد معين من القيم المتقطعة "المنفصلة". وهكذا تصبح الإشارة بشكلها الجديد جاهزة للتشفير بالطريقة المناسبة.

يمتد طيف الإشارة الكلامية حتى 12 كيلوهرتز [2]، وبالاعتماد على معيار "شانون" يكون تواتر التجزئة إلى عينات مساوياً 24 كيلوهرتز. فعند التعرف الآلي على الكلام أو التركيب الاصطناعي له، فإن التواتر يمتد من 6 إلى 16 كيلوهرتز. ولكن في ميدان الاتصالات يتم حصر قيمة الطيف بهدف تقليص تكاليف المعالجة الرقمية.

ففي مجال الاتصال الهاتفي، تعتبر نوعية الكلام كافية باختيار طيف لا يتعدى 3400 هرتز، وبالتالي فإن تواتر التجزئة المتبني هو 8 كيلوهرتز. فالعينات المقسمة كمياً تشفر باستعمال كلمات مؤلفة من 8 أرقام ثنائية. فيكون التدفق الثنائي الناتج هو 64 كيلورقم ثنائي/ ثانية. وتدعى هذه العملية التغيير بالتبضات المشفرة.

تتيح خوارزميات تشفير الكلام التي تستعمل نظام محاكاة جهاز النطق (حنجرة، حبال صوتية، ...) الحصول على تدفقات متدنية تصل إلى حدود 2 كيلورقم ثنائي/ ثانية، وهي نتيجة جد مرضية لو لم تكن النوعية رديئة، وهذا يرجع -كما رأينا سابقاً- إلى التشوه الكبير.

إن الهدف الأساسي من وراء تقنيات التشفير العديدة التي أحدثت إلى يومنا هذا هو الوصول إلى نظام يوفق بين الكفاءة والتوعية والتدفق الثنائي والتكاليف.

1.3. تعريف مبسط للرقمنة

إن الإشارة الكلامية بعد تسجيلها تتحول إلى أخرى كهربائية تكون بموجب ذلك مستمرة، ولتوضيح عملية رقميتها فضلنا تقديم تعريفات مبسطة مدعومة بأشكال بيانية.

❖ التعريف الأول

تحتوي الإشارة المستمرة على قيم مستمرة للسعة وللزمن مثلما هو مبين في الشكل (1.3).

❖ التعريف الثاني

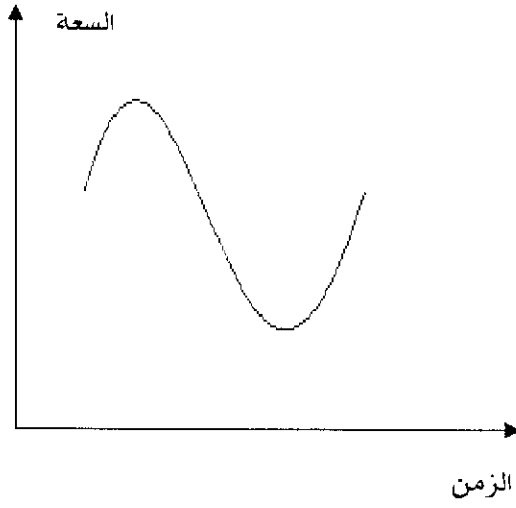
يمثل الشكل (2.3) إشارة ذات سعة مقسمة كمياً، أما الزمن فهو مستمر.

❖ التعريف الثالث

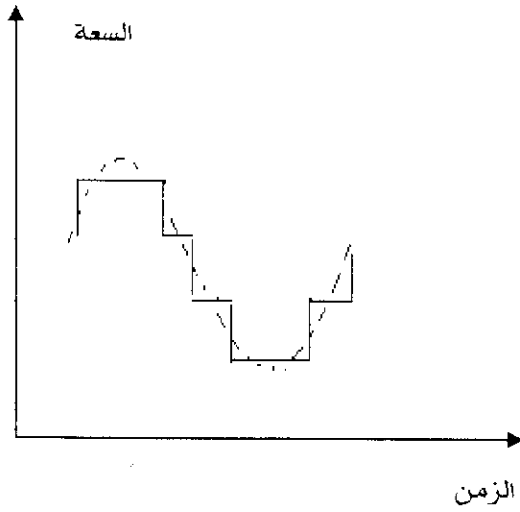
تدعى الإشارة ذات السعة المستمرة والزمن المتقطع بالإشارة المجزأة إلى عينات، ونجدها ممثلة بالشكل (3.3).

❖ التعريف الرابع

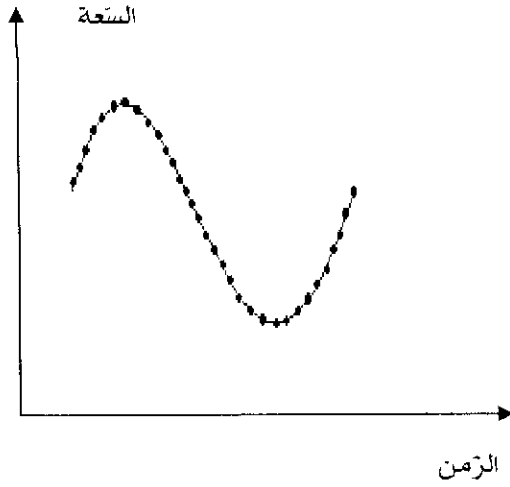
إن الإشارة ذات السعة والزمن المتقطعين، المبينة في الشكل (4.3) هي إشارة مرقمنة.



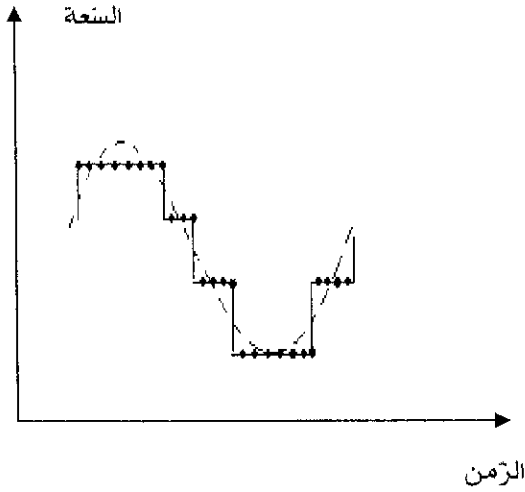
الشكل (1.3). إشارة مستمرة



الشكل (2.3). التقسيم الكمي للسعة



الشكل (3.3). إشارة مجزأة إلى عينات



الشكل (4.3). إشارة مرقمنة

4. قياس نوعية المشفر

1.4. القياس الموضوعي

إن القياس الموضوعي المعمول به لتقييم نوعية المشفر هو حاصل قسمة الإشارة بالنسبة للضجيج ويعطى بالمعادلة التالية :

$$SNR = 10 \log \frac{\sigma_s}{\sigma_n} \quad (4.1)$$

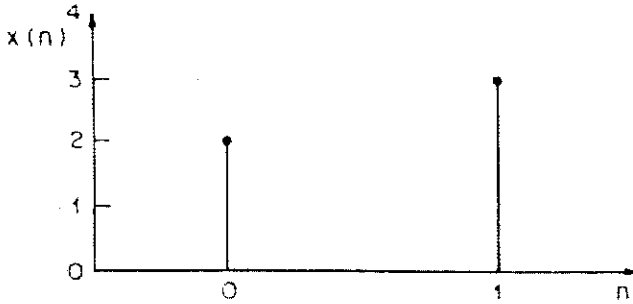
إلا أن هذا المفهوم أصبح غير فعال تجريبياً لقياس حسية مشفر ذي نوعية جيدة، لذلك أدخل مفهوم الحاصل الجزئي لقسمة الإشارة بالنسبة للضجيج وهو عبارة عن متوسط قيم الحواصل الجزئية لكل نافذة من الإشارة. كما توجد معايير أخرى للتقييم مثل مسافة "إيتاكورا" الموزونة.

2.4. القياس الذاتي

يعتبر الاعتماد على الأذن البشرية شيئاً ضرورياً لتغطية النقص التاجم عن القياس الموضوعي بحيث لا يستطيع هذا الأخير أن يوفر تقييماً دقيقاً في بعض الحالات لأنه لا يأخذ في الاعتبار الخصائص الحسية التي تملكها الأذن.

5. مفهوم التحويل

لقد أردنا من خلال المثال الذي نحن بصدد عرضه في هذا الفصل إعطاء تفسير مبسط لمفهوم التحويل [3].
لتكن الإشارة $\{x(n)\}$ الممثلة في الرسم البياني (5.1) :



الشكل (5.1). الإشارة $x(n)$ [3]

كما نلاحظ في الشكل (5.1) فرتبة هذا التحويل هي 2.

$$x(0)=2 ; x(1)=3 \quad (5.1)$$

نستطيع كتابة الإشارة $\{x(n)\}$ على الشكل التقريبي الممثل لها كالآتي :

$$\{5/2, 5/2\} = \frac{5}{\sqrt{2}} \left\{ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\} = \frac{5}{\sqrt{2}} \{b_{00}, b_{01}\} \quad (5.2)$$

يمكن وصف الإشارة $\{x(n)\}$ بدقة بإضافة القيمة $\{-0.5, +0.5\}$ للعلاقة (5.2).

$$\{-0.5, +0.5\} = \frac{-1}{\sqrt{2}} \left\{ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\} = \frac{-1}{\sqrt{2}} \{b_{10}, b_{11}\} \quad (5.3)$$

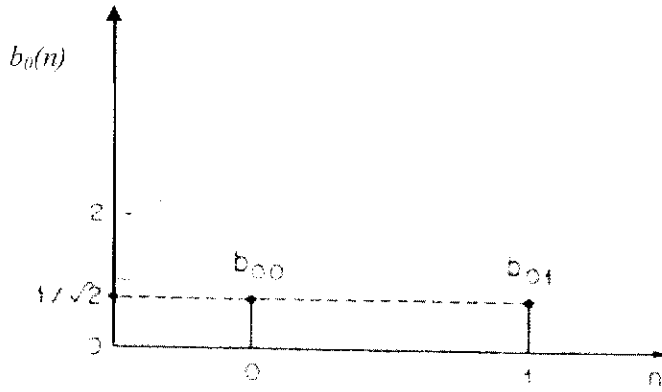
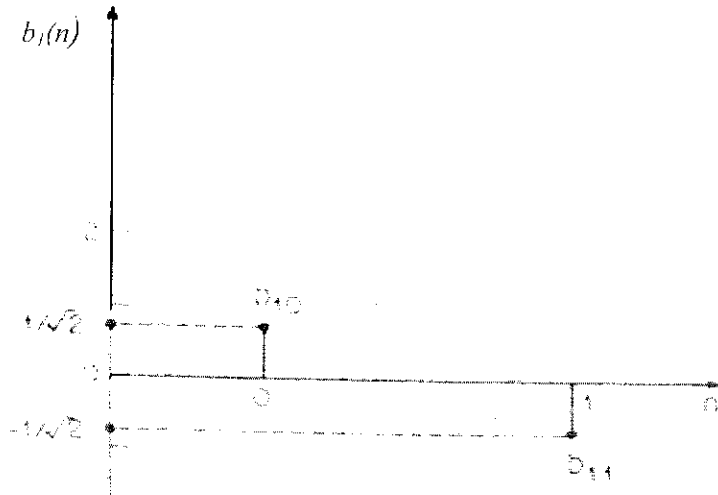
يعتبر الشعاعان b_0 و b_1 المعرفان في العلاقتين (5.3) و (5.2) والموضحين في الشكلين (5.3) و (5.2) شعاعي قاعدة، إذ يمكن باستعمالهما تمثيل الإشارة $\{x(n)\}$ بتنسيق خطي موزون بمعاملات التحويل.

$$T(0) = \frac{5}{\sqrt{2}} = \frac{1}{\sqrt{2}} (x(0) + x(1)) \quad (5.4)$$

$$T(1) = \frac{-1}{\sqrt{2}} = \frac{1}{\sqrt{2}} (x(0) - x(1)) \quad (5.5)$$

يمثل المعاملان المحولان $T(1)$ و $T(0)$ على الترتيب مجموع عيني الإشارة $x(n)$ والفرق بينهما. ويمكن القول بعبارة أخرى أنهما يمثلان الطاقة المتعلقة بالمركبات منخفضة التواتر والأخرى ذات التواتر العالي في الإشارة $x(n)$.

وعلى هذا الأساس نستنتج أن عزل المركبات المنخفضة عن تلك العالية هو الدافع وراء التمثيل باستعمال التحويل [3].

الشكل (5.2) الشعاع القاعدي b_0 [3]الشكل (5.3) الشعاع القاعدي b_1 [3]

باستطاعتنا استرجاع القيم $x(0)$ و $x(1)$ بالعمليتين العكسيتين التاليتين :

$$x(0) = \frac{1}{\sqrt{2}}(T(0) + T(1)) \quad (5.6)$$

$$x(1) = \frac{1}{\sqrt{2}}(T(0) - T(1)) \quad (5.7)$$

بصفة عامة تعطي عملية التحويل معاملات غير مترابطة، بشروط تغيير مختلفة. ونستطيع ببسر استنتاج العبارة (5.8) انطلاقاً من الإشارة العشوائية المستقرة $x(n)$ والمحوّلة بالمعادلتين (5.5) و (5.4) :

$$E(T(0)T(1)) = E[x(0)^2] - E[x(1)^2] = 0 \quad (5.8)$$

1.5. تفسير

لقد قمنا بتمثيل فكّ الترابط الناتج عن عملية التحويل تمثيلاً هندسياً وهذا بواسطة الشكل (5.4).

باعتبار الإشارة $x(n)$ نقطة في مجال ذي بعدين (x_0, x_1) نلاحظ أنّ الشعاعين b_0 و b_1 متعامدان بحيث يشكلان مجالاً جديداً (T_0, T_1) تكون بموجبه النقطة ممثلة بالإحداثيات $(T(0), T(1))$.

6. التشفير باستعمال التحويل

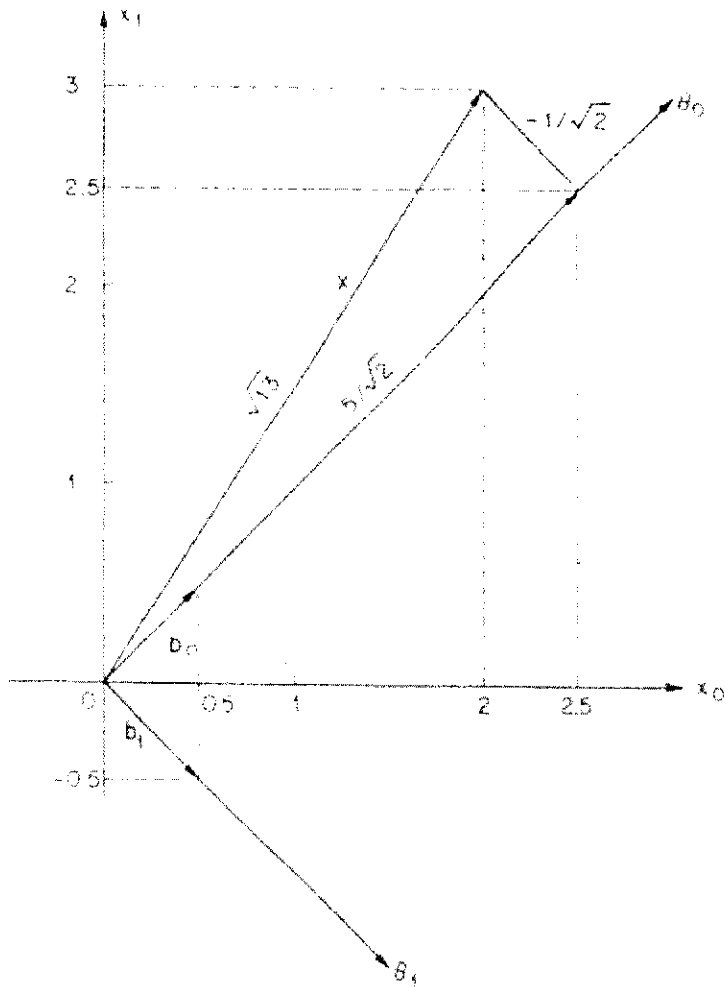
تقتضي تقنية الضغط بالمحولات معالجة الإشارة الرقمية عن طريق التحويلات المتعامدة ثم بعد ذلك تشفير الإشارة الناتجة لإنقاص كمية المعلومات [4].

ويمكن لنا سرد بعض التحويلات على سبيل المثال : SVD، المحولة الجيبية DCT ومحولة "فوريي" [5].

بعبارة أدق، يتم التشفير باستعمال المحولات عن طريق تطبيق تحويل خطي على مجموعة من العينات المترابطة المكونة لإشارة المدخل بغية الحصول على مجموعة من العوامل غير المترابطة، بحيث يخضع جزء منها إلى التقسيم الكمي ثم إلى التشفير. ومعلوم أن هذا الجزء هو الذي يحتوي على الكمية الكبرى من المعلومات.

إن أغلب التحويلات تقوم بإلغاء العوامل ذات الطاقة الدنيا، فمن المستحسن إذا استعمال المحولة التي تقوم بتخزين أكبر كمية من الطاقة في عدد قليل من العوامل بحيث يتسنى بعد ذلك اختيار تلك التي يجب إرسالها.

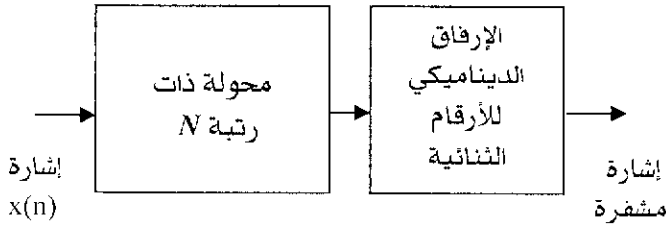
توجد عدة طرق لتحقيق ذلك، من بينها طريقة تعتمد على وضع قيمة معينة يتم بموجبها إلغاء العوامل ذات السعة التي لم تبلغ هذه القيمة وباختيار العوامل الأخرى [4].



الشكل (5.4) التمثيل الهندسي لتحويل المدخل (x_0, x_1) [3]

كما توجد طريقة أخرى تعتمد على حساب شرط التغير الخاص بكل مركبة تحويل، ومن ثم تطبيق خوارزمية الإرفاق الديناميكي للأرقام الثنائية للحصول على تقسيم كمي دقيق ونوعي لبعض العوامل دون غيرها [6.5].

تتعلق فعالية نظام التشفير بالتحويل بنوع المحولة وكذلك بإرفاق الأرقام الثنائية أثناء التشفير [6]. انظر الشكل (6.1) .



الشكل (6.1) التشفير "الضغط" بالتحويل.

1.6. الإرفاق الديناميكي للأرقام الثنائية

يعتبر الإرفاق الديناميكي للأرقام الثنائية جزءاً مهماً في التشفير باستعمال التحويل، إذ يعتمد أساساً على جعل القيمة الرباعية المتوسطة لخطأ الاسترجاع متدنية. ويسمح بالتقسيم الكمي الدقيق لبعض المركبات دون غيرها، مما يتيح كسباً إضافياً معتبراً بالمقارنة مع طرق التشفير التقليدية.

إن قيم شروط التغيير لركبات التحويل مختلفة وبالتالي فتدفعاتها R_k مختلفة. ويكون التدفق المتوسط R ثابتا بحيث يعبر عنه بالعلاقة (6.1) :

$$R = \frac{1}{N} \sum_{k=0}^{N-1} R_k = \text{constante} \quad (6.1)$$

أما الإرفاق الأمثل للأرقام الثنائية فيعتمد على توزيع شروط التغيير لركبات التحويل [5] ويعبر عنه كالآتي :

$$R_k = R + \frac{1}{2} \log_2 \frac{\sigma_k^2}{\left[\prod_{j=0}^{N-1} \sigma_j^2 \right]^{\frac{1}{N}}} \quad (6.2)$$

7. التحويل SVD

يعتبر التحليل إلى القيم الشاذة SVD من أهم الأدوات الأساسية في الجبر الخطي، واستعماله الآن يجد رواجا متزايدا في التطبيقات المختلفة لمعالجة الإشارة وكذا لحل بعض المشاكل المتعلقة بالتحليل العددي. ويمكن تعريفها بأنها تعميم للتمثيل المعروف بانقيم والأشعة الصرفة لمصفوفة ما [7].

1.7. القيم والأشعة الصرفة

لتكن مجموعة تتكون من n معادلة خطية من النوع التالي :

$$AX = \lambda X \quad \text{ou} \quad a_{ij}x^j = \lambda x_i \quad (7.1)$$

A هي مصفوفة مربعة و X عمود مصفوف. تكتب المعادلة (7.1) على الشكل التالي :

$$(A - \lambda U)X = 0 \quad \text{ou} \quad (a_{ij} - \lambda \delta_{ij})x^j = 0 \quad (7.2)$$

بجيث U هي مصفوفة وحدة ذات رتبة n .
الحل $X=0$ لا يمثل أي فائدة. لكي يقبل النظام حولا أخرى غير معدومة يجب أن يكون المحدد معدوما.

$$\|A - \lambda U\| = 0 \quad \text{ou} \quad \|a_{ij} - \lambda \delta_{ij}\| = 0 \quad (7.3)$$

$$\begin{vmatrix} (a_{11} - \lambda) & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & (a_{22} - \lambda) & & \dots & a_{2n} \\ a_{31} & & (a_{33} - \lambda) & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & (a_{nn} - \lambda) \end{vmatrix} = 0 \quad (7.4)$$

بنشر المحدد نحصل على معادلة من الدرجة n تسمى المعادلة المميزة للمصفوفة A وتكتب كالاتي :

$$(-1)^n \cdot [\lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \dots + a_{n-1} \lambda + a_n] = 0 \quad (7.5)$$

كما تعتبر الجذور λ_i لهذه المعادلة وعددها n ، حلولا غير معدومة للعبارة (7.1) بحيث تمثل القيم الصرفة للمصفوفة A .
كل قيمة صرفة λ_i يقابلها حل غير معدوم X_i لـ (7.1) يدعى بالشعاع الصرف للمصفوفة A والرفق λ_i .

2.7. التحليل إلى قيم شاذة لمصفوفة

كل مصفوفة A ذات حجم $M \times N$ بحيث $M \geq N$ ، تقبل التمثيل التالي :

$$A = USV^T; \quad (7.6)$$

تعتبر S مصفوفة قطرية ذات حجم $N \times N$.

$$S = \text{diag}(s_0, s_1, \dots, s_{N-1}) \quad (7.7)$$

إن العناصر القطرية لهذه المصفوفة هي عناصر حقيقية موجبة ومرتبة ترتيبا تنازليا كما هو موضح في العبارة (7.8).

$$S_0 \geq S_1 \geq \dots \geq S_{N-1} > 0 \quad (7.8)$$

بينما U و V هما مصفوفتان موحدتان بحجم $N \times N$:

$$U = (U_0, U_1, \dots, U_N) \quad (7.9)$$

$$V = (V_0, V_1, \dots, V_{N-1}) \quad (7.10)$$

$$U U^T = U^T U = I_N; \quad U^T = U^{-1} \quad (7.11)$$

$$V V^T = V^T V = I_N; \quad V^T = V^{-1} \quad (7.12)$$

التحليل إلى القيم الشاذة للمصفوفة A يمكن كتابته على الشكل التالي :

$$A = \sum_{i=0}^{N-1} S_i U_i V_i^T \quad (7.13)$$

تشكل الأشعة الصرفة للمصفوفة H H^T أعمدة المصفوفة الموحدة

V ، أما الأشعة الصرفة للمصفوفة $H H^T$ فهي أعمدة المصفوفة U .

ويعتبر مربع القيم S_i للمصفوفة S مساويا للقيم الصرفة للمصفوفتين U و V .

3.7. تفسير التحويل SVD

لقد رأينا سابقا بأن التحويل SVD لمصفوفة A ذات حجم $N \times N$ معرف

بالعبارة (7.13) التي يمكن نشرها حسب المعادلة (7.14).

$$\begin{aligned} [a_0, a_1, \dots, a_{N-1}] &= S_0 U_0 V_0^T + \dots + S_{N-1} U_{N-1} V_{N-1}^T \\ &= S_0 [V_{00} U_0, \dots, V_{0N-1} U_0] + S_1 [V_{11} U_1, \dots, V_{1N-1} U_1] + \\ &\quad \dots + S_{N-1} [V_{N-11} U_{N-1}, \dots, V_{N-1N-1} U_{N-1}] \end{aligned} \quad (7.14)$$

بملاحظة كل عمود للمصفوفة A نستطيع أن نستنتج ببسر العبارة :

$$a_i = \sum_{j=0}^{N-1} S_j V_{ji} U_j \quad i=1,2,\dots,N-1. \quad (7.15)$$

بعبارة أوضح، فإن حساب التحويل SVD للمصفوفة A يرجع أساسا إلى كتابة كل عمود منها على شكل مجموع موزون للأشعة القاعدية U_j بحيث $(j=0, \dots, N-1)$.

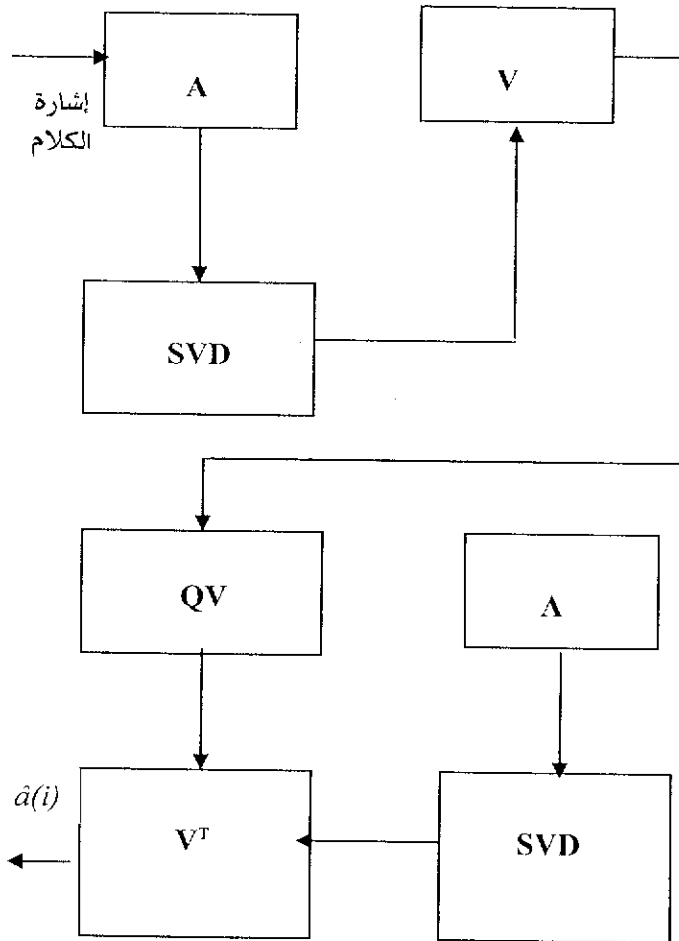
إن معاملات الموازنة للعمود a_i هي حاصل ضرب القيم S_j ومركبات العمود ذي الرتبة i للشعاع V_j .

8. التشفير باستعمال SVD

يلعب التحويل SVD دورا معتبرا في معالجة الإشارة، بحيث له القدرة على استخراج فضاء العطيات المندمج في الضجيج [6].

وبخصوص إشارة الكلام، يقوم بفك ترابط عيناتها بعزل الأصوات المجهورة عن تلك غير المجهورة (الإشارات ذات الطاقة العالية عن تلك ذات الطاقة المنخفضة).

إن نظام التشفير باستعمال SVD يتطلب تطبيق هذه الأخيرة على المصفوفة A التي تمثل معاملات التنبؤ الخطي $a(i)$ ومن ثم استخراج المصفوفة V التي تعتبر مصفوفة التحويل المستعملة في تجربتنا. إنه من الضروري أن تكون هذه المصفوفة متعامدة وموحدة في نظام التشفير بالتحويل [5]. ومن ثم يتم تطبيق التقسيم الكمي الشعاعي على المعاملات المحولة (عناصر المصفوفة V).



الشكل (8.1) نظام التشفير باستعمال التحويل SVD

عند الاستقبال، يتم استعادة معاملات التنبؤ $\hat{a}(i)$ انطلاقاً من المعاملات المحولة المرسل، والمعطاة بالعلاقة التالية :

$$A = USV^T$$

بحيث تحتوي المصفوفة A على المعاملات $\hat{a}(i)$ المستعادة. ويمثل الشكل البياني (8.1) نظام التشفير باستعمال المحولة SVD.

9. التقسيم الكمي الشعاعي

نقدم، في هذا الفصل، تعريفاً للتقسيم الكمي الشعاعي مع تفصيل لإحدى الخوارزميات المستعملة لهذا الغرض، وهي طريقة LBG، عارضين النتائج الدالة على فعاليتها.

إن التقسيم الكمي الشعاعي، كما يدل عليه اسمه، هو التقسيم الكمي لشعاع ما [8]. وهو تقنية مستعملة في ضغط الإشارات في الآونة الأخيرة، سواء تعلق الأمر بإشارة الكلام أو بالصورة

1.9. تعريف

يعتبر المقسم الكمي الشعاعي Q ذو البعد k والحجم N تطبيقاً من شعاع ينتمي إلى فضاء إقليدي E_k ذي K بعد، نحو مجموعة منتهية C محتوية على N شعاع. وتدعى هذه الأشعة بأشعة الشفرة.

$$Q : E_k \rightarrow C$$

مع $C = \{y_1, y_2, y_3, \dots, y_N\}$ و $y_i \in E_k$ من أجل كل قيمة i تنتمي إلى $\{1, 2, \dots, N\}$.

هناك نوعان للتقسيم الكمي الشعاعي (ت.ك.ش) : جبري و إحصائي.
 يتيح (ت.ك.ش) تمثيل شعاع x ذي k مركبات بشعاع آخر y ينتمي إلى مجموعة منتهية تتكون من M شعاع. تسمى هذه المجموعة بالقاموس.
 إن تعويض x بالشعاع y يؤدي إلى حدوث تشوه أو مسافة يرمز لها بالقيمة $d(x,y)$. هناك طرق عدة لحساب هذا التشوه، نذكر منها على سبيل المثال المسافة الطيفية ومسافة الطيف العكسي ومسافة "إيتاكورا-سايتو". وهناك مسافة أخرى تستعمل بكثرة هي المسافة الإقليدية المثلة بالعلاقة (9.1).

$$d_2(x,y) = \frac{1}{k} \sum_{i=1}^k (x_i - y_i)^2 \quad (9.1)$$

إن الهدف من وراء التقسيم الكمي الشعاعي هو جعل قيمة التشوه المتوسط دنيا، باعتبار عدد معين من الأقسام. أو تقليص عدد هذه الأقسام بتحديد قيمة معينة للتشوه المتوسط.

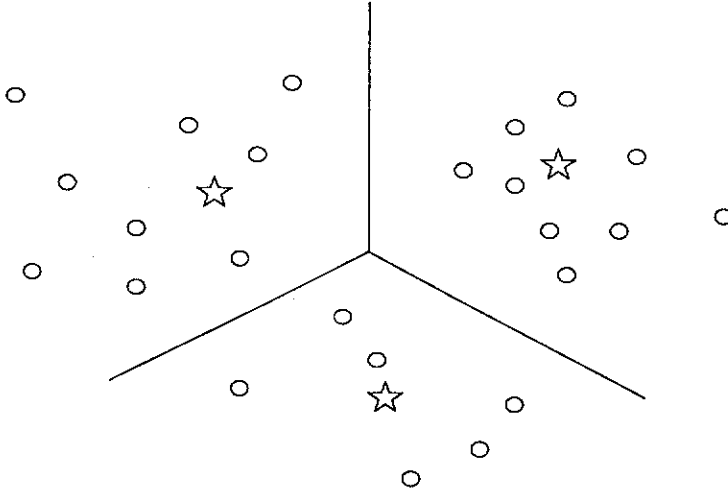
من أجل هذا الهدف يتم اللجوء عادة إلى طرق عدة منها : خوارزمية "لويد" وخوارزمية LBG التي سيتم التطرق إليها لاحقا.

2.9. التقسيم الكمي الشعاعي الإحصائي

هناك نوعان من التصنيف الآلي، تدرجي وغير تدرجي. والأول في حقيقة الأمر قليل الاستعمال في المعالجة الآلية للكلام.

ويعتبر التقسيم الكمي الشعاعي الإحصائي طريقة غير تدرجية تؤدي إلى تجزئة وحيدة للمعطيات إلى أقسام : إذ تقسم مجموعة التمرن

إلى مجموعات منفصلة تمثل كل واحدة منها بمركز جاذبيتها، والذي يساهم من خلال الحسابات إلى إيجاد التجزئة النهائية. من الشكل (9.1) نستنتج بأن كل قسم، يعرف بأنه مجموعة النقاط القريبة من مركز جاذبيتها والبعيدة عن المراكز الأخرى.



الشكل (9.1) توزيع الأقسام

3.9. التقسيم الكمي الشعاعي الجبري

تعتمد هذه الطريقة على استعمال قاموس يملك بنية رياضية قوية يمكنها من الاستفادة من البنى الجبرية المنتظمة مثل الشفرات وشبكات النقاط لتصميم "مقسّمات كمية" فعالة، غير معقدة ولا تحتاج للتخزين.

4.9. خوارزمية LBG

لنعتبر مجموعة تتكون من عدد L من الأشعة x نريد توزيعها إلى عدد M من الأقسام محدد مسبقا [9].

نلخص مراحل الخوارزمية في النقاط التالية :

- اختيار عدد M من مراكز الجاذبية y_i بصفة عشوائية في الفضاء E_k .
- إرفاق كل شعاع x إلى قسمه بحساب المسافات بين هذا الشعاع وكل مراكز الجاذبية الموجودة. بحيث يمثل الشعاع بالمركز الأقرب منه.
- حساب موقع كل مركز y_i مجددا حسب التموقع الحالي للأشعة لإنقاص قيمة المسافة D_i بحيث :

$$D_i = \sum_j d(x_j^{(i)}, y_i) \quad (9.2)$$

تمثل القيمة $d(x_j^{(i)}, y_i)$ المسافة بين $x_j^{(i)}$ و y_i . أما $x_j^{(i)}$ فهو مجموعة الأشعة التي تنتمي إلى القسم i .

- حساب المسافة الكلية D :

$$D = \sum_{i=1}^M D_i \quad (9.3)$$

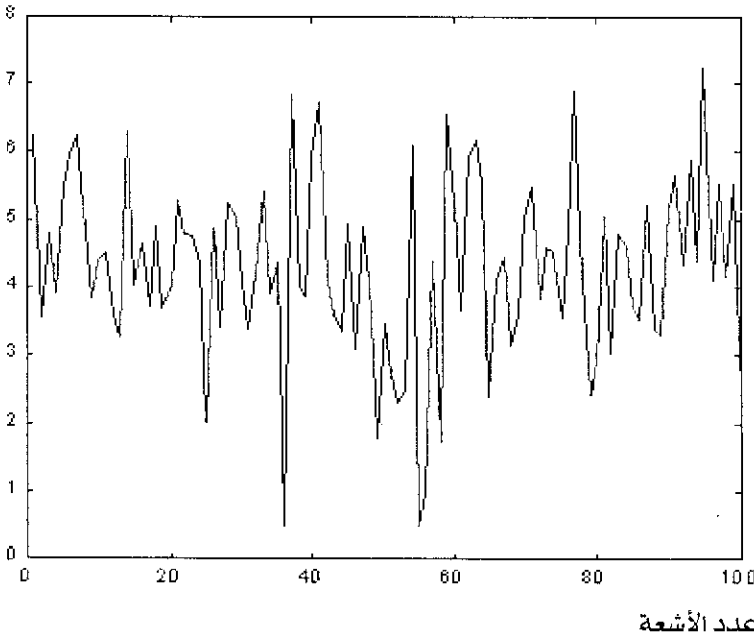
- تكرار المراحل السابقة حتى تصل المسافة D إلى حد تتغير فيه بأقل من ϵ وهي قيمة يتم اختيارها للحصول على الدقة المطلوبة.

عند استعمال هذه الطريقة، يتم اختيار مراكز الجاذبية بصفة عشوائية في البداية، مما يدل على إمكانية حدوث اختيار يؤدي إلى مسافة أصغر. ولهذا ينصح باختبار عدة اختيارات.

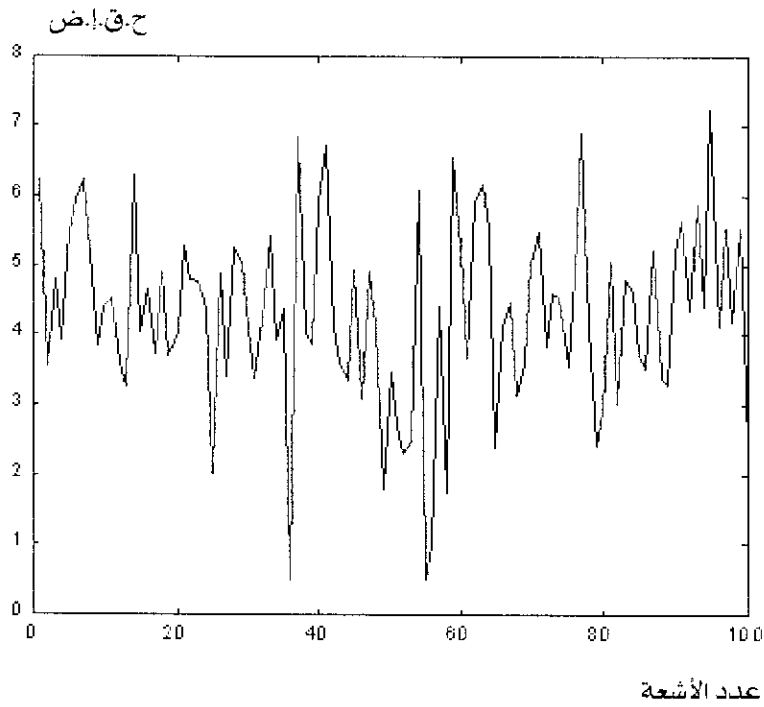
5.9. كفاءة المقسم الكمي الشعاعي

لكي نظهر كفاءة المقسم الكمي قمنا بتوليد إشارة "غوسية" وجعلناها في مدخله من أجل إخضاعها لعملية "ت. ك. ش" باستعمال أحجام مختلفة للقاموس. والجدير بالذكر أن كل شعاع من الإشارة "الغوسية" يحتوي على عشر مركبات. وتعبّر الأشكال الآتية عن النتائج المحصل عليها باستخدام أحجام القاموس المتعددة، بحيث نلاحظ ارتفاع قيمة حاصل قسمة الإشارة بالنسبة إلى الضجيج "ح. ق. إ. ض" كلما كان حجم القاموس كبيرا، ويرى في الجدول (9.1) بشكل واضح. (الصفحة 44).

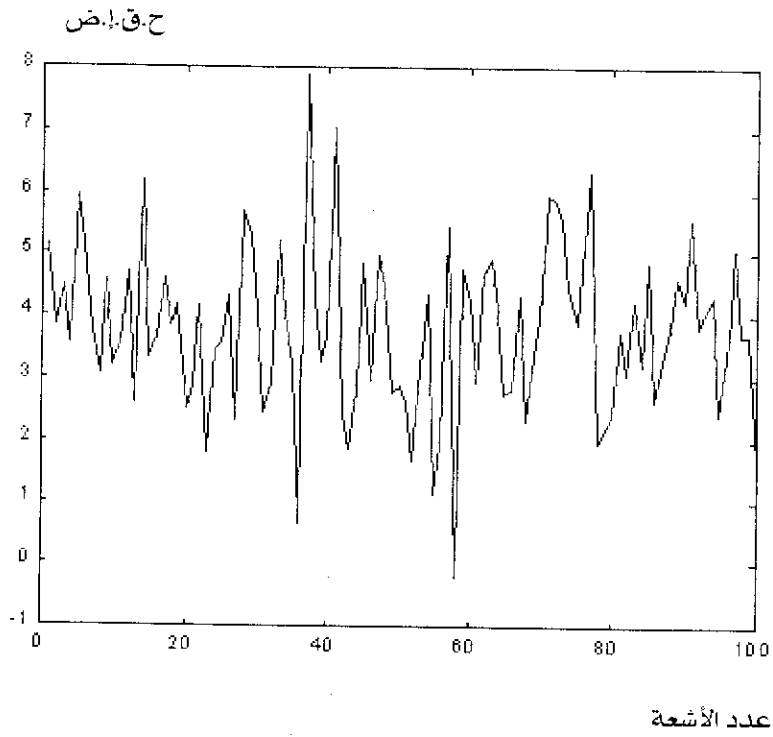
ح.ق.إ.ض



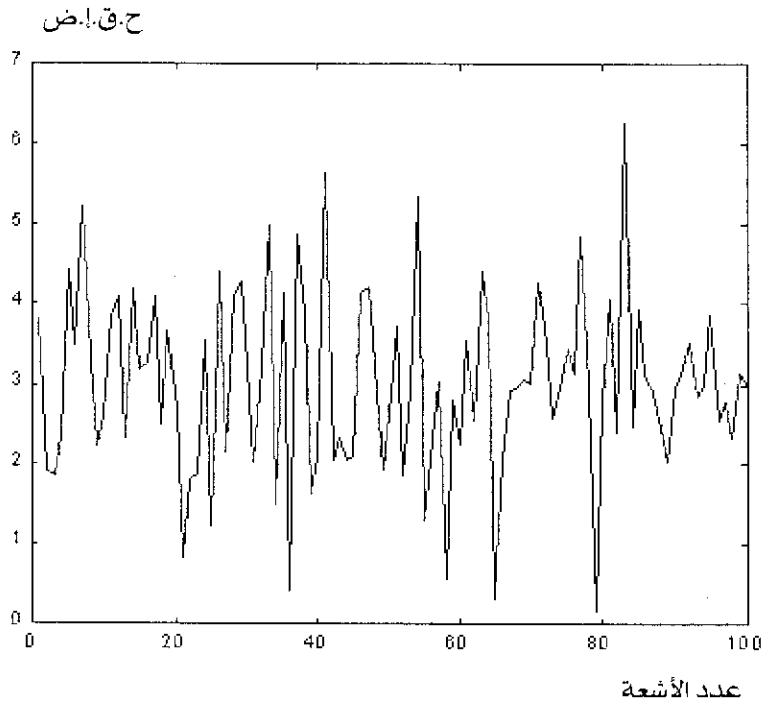
الشكل (9.2) حاصل قسمة الإشارة بالنسبة للضجيج في حالة قاموس ذي حجم 512 شعاع



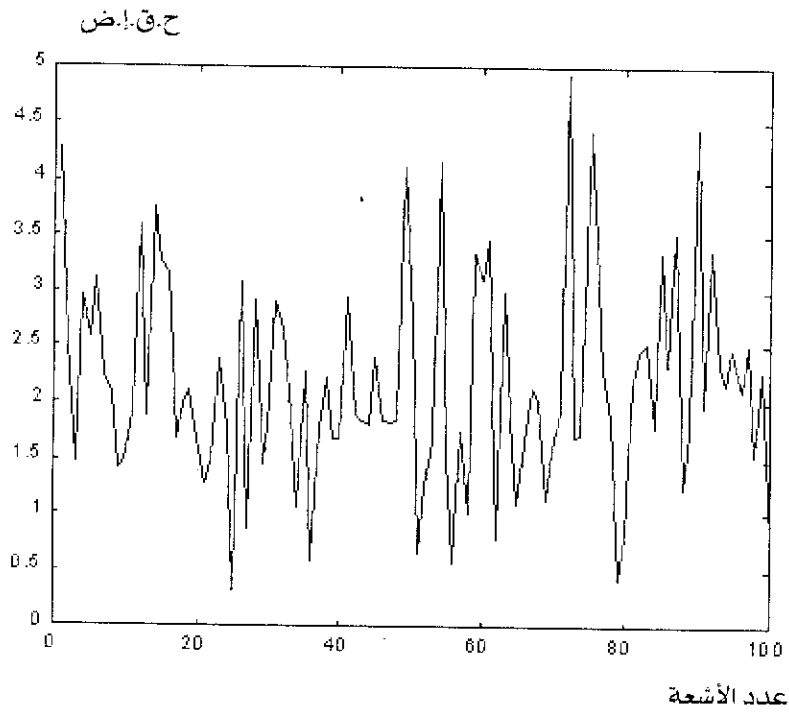
الشكل (9.3) حاصل قسمة الإشارة بالنسبة للضجيج
في حالة قاموس ذي حجم 256 شعاع



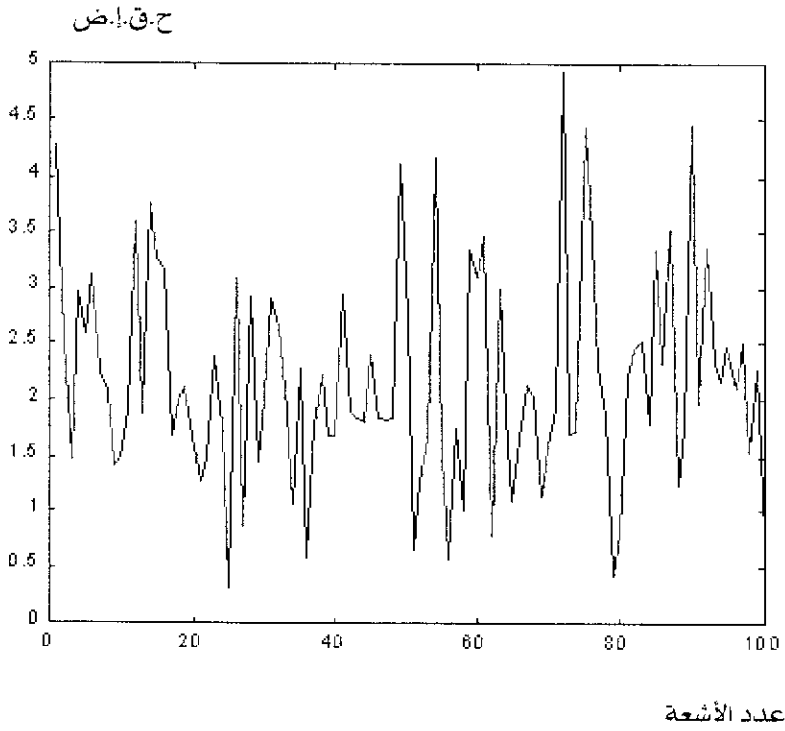
الشكل (9.4) حاصل قسمة الإشارة بالنسبة للضجيج
في حالة قاموس ذي حجم 128 شعاع



الشكل (9.5) حاصل قسمة الإشارة بالنسبة للضجيج
في حالة قاموس ذي حجم 64 شعاع



الشكل (9.6) حاصل قسمة الإشارة بالنسبة للضجيج
في حالة قاموس ذي حجم 32 شعاع



الشكل (9.7) حاصل قسمة الإشارة بالنسبة للضجيج
في حالة قاموس ذي حجم 16 شعاع

512	256	128	64	32	16	حجم القاموس
4.31	4.31	3.78	2.99	2.17	1.66	ح.ق.إ.ض (dB)

9.1 . جدول ممثل لقيم "ح.ق.إ.ض" الناتجة عن أحجام مختلفة للقاموس.

10. تجارب ونتائج

نقدم في هذا الفصل النتائج المتحصل عليها عن طريق التجارب التي قمنا بها والمتعلقة بتشفير "ضغط" الكلام باستعمال المحولة SVD كما هو مبين في الشكل (8.1).

إن معاملات المدخل لنظام التشفير هي معاملات التنبؤ الخطي $a(i)$ والتي استخرجنا منها عشرة عناصر لكل نافذة من الإشارة الكلامية. وهي تمثل عناصر المصفوفة A التي طبقنا عليها التحويل إلى القيم الشاذة. إن المصفوفة V هي مصفوفة التحويل المستعملة.

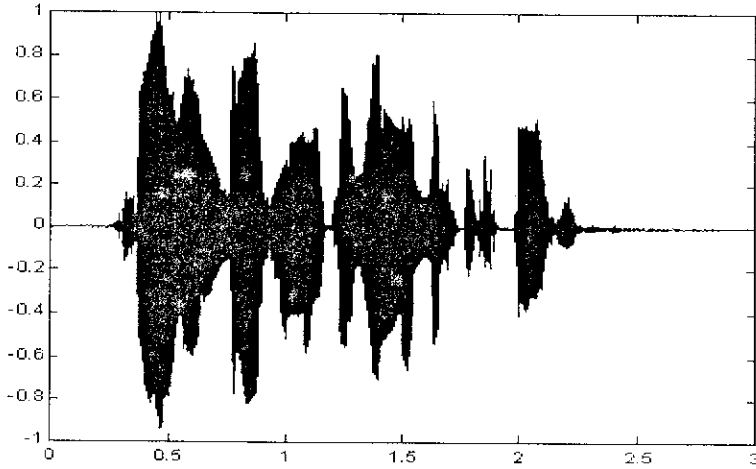
بعد ذلك قمنا بتطبيق التقسيم الكمي الشعاعي على معاملات المصفوفة V باستعمال قاموس للأشعة المثلة ذي حجم 128 كنا قد كوناه بواسطة خوارزمية LBG.

ولاسترجاع المعاملات $\hat{a}(i)$ استعملنا العلاقة :

$$A = USV^T$$

وبتحريض مرشح الإصدار ذي المعاملات $\hat{a}(i)$ بالإشارة المتبقية تم الحصول على الإشارة المشفرة.

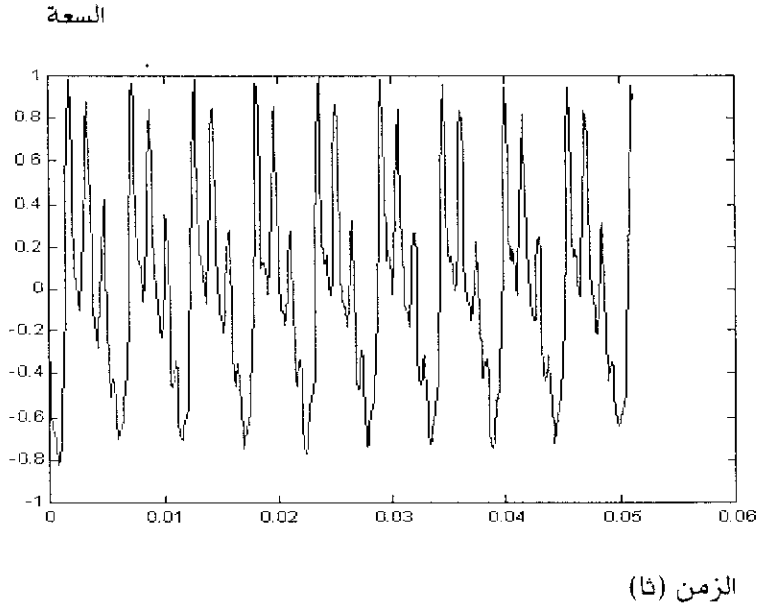
السعة



الزمن (ثا)

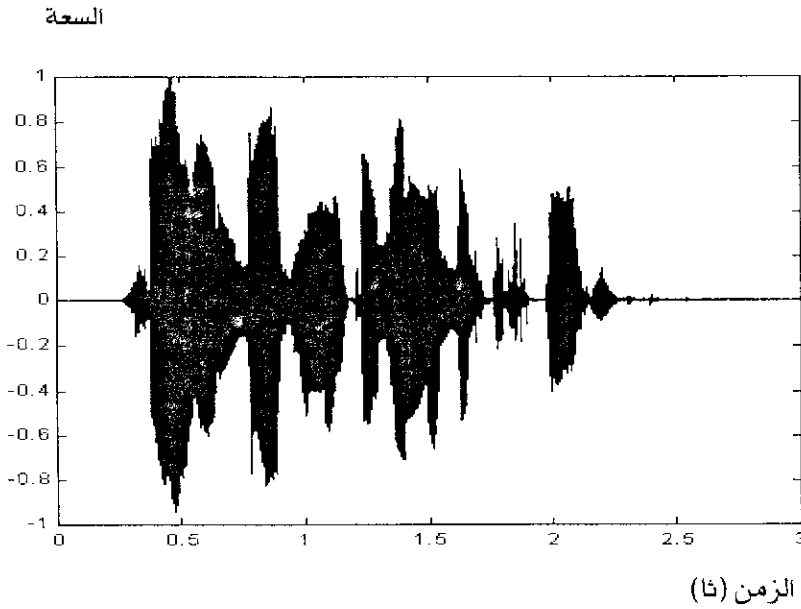
الشكل (10.1) إشارة الكلام المتعلقة بالجملة قبل عملية التشفير.

لقد استعملنا جملا عديدة للاختبار، ونظرا للنتائج التقاربة المتحصل عليها اکتفينا بعرض تلك المتعلقة بجملة لا تتعدى مدتها ثلاث ثوان - ممثلة بالشكل (10.1) - وكذلك مقطوعة صغيرة من الإشارة الكلامية متعلقة بحركة الفتحة - الشكل (10.2) -.



الشكل (10.2) إشارة الكلام المتعلقة بحركة الفتحة قبل عملية التشفير.

يوضح الشكل (10.3) الموجة الزمنية المستعادة للجملة. أما الشكل (10.4) فيمثل تلك المتعلقة بحركة الفتحة. بمقارنتهما مع الإشارتين الأصليتين المبينتين في الشكلين (10.1) و(10.2) نستطيع ملاحظة التشابه الكبير الموجود بينهما.

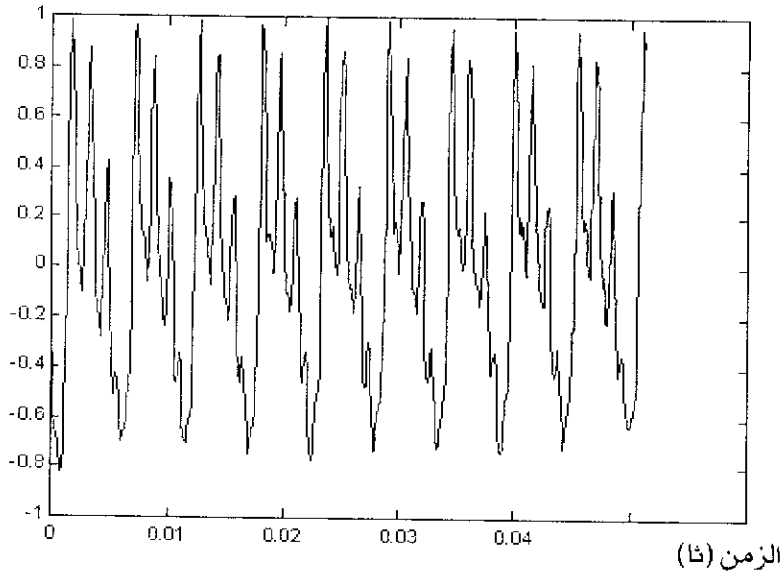


الشكل (10.3) إشارة الكلام المتعلقة بالجملة
بعد عملية التشفير.

إلا أن هذا التشابه لا يعني بالضرورة أن هناك نوعية جيدة لإشارة الكلام المشفرة [11، 10]. ولذلك وجب استعمال القياس التالي.

إحدى هذه القياسات الناتية والتي تعرف استعمالا واسعا هو "قياس الرأي المتوسط" حيث تقتضي سماع الإشارات من طرف عدد من الأشخاص يتراوح بين 12 و 24. يقومون بعد ذلك بوضع نقاط حسب التشوهات التي تصيب الإشارة والتي تلتقطها الأذن.

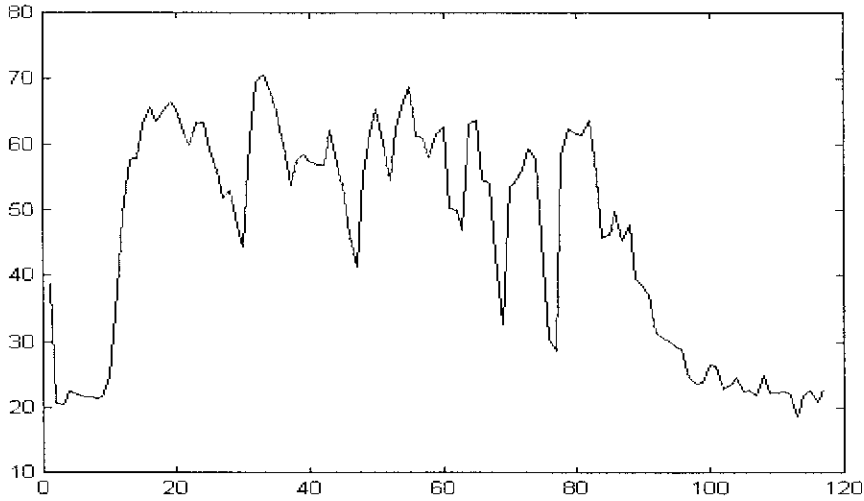
السعة



الشكل (10.4) الموجة الزمنية المتعلقة بحركة الفتحة
بعد عملية التشفير.

يتراوح سلم التنقيط الذي يعتمد عليه القياس التالي بين القيمتين 1 و5، بحيث تعبر النقطة 5 عن عدم وجود تشوهات محسوسة، بينما تدل العلامة 1 عن وجود ضجيج مزعج. فبالنسبة لتجارينا أعطيت العلامة 5 لنوعية الإشارات المستعادة.

ح.ق.إ.ض



عدد الأشعة

الشكل (10.5) حاصل قسمة الإشارة بالنسبة للضجيج
بالنسبة للجملة

يؤكد الشكل (10.5) التشابه الظاهر بين الإشارات الأصلية
وتلك المستعادة بعد عملية التشفير، وهذا بملاحظة المنحنى
الخاص بقيم "ح.ق.إ.ض".

إن متوسط قيم "ح.ق.إ.ض" لكل الإشارات المستعملة في تجاربنا
يضاهي القيمة 46 dB.

11. الخاتمة

أفضى الاستماع إلى الإشارات الكلامية المستعادة إلى الإقرار بوجود نوعية مُرضية، خالية من أي ضجيج. وقد جاء هذا الإقرار نتيجة لتقييم قام به العديد من الأشخاص.

يهدف اللجوء إلى التشفير باستعمال التحويل إلى التقليل من قيمة التدفق، وهذا يتم، كما هو معلوم، بفضل تركّز طاقة الإشارة في عدد ضئيل من معاملات التحويل. لهذا الغرض يستعمل التحليل إلى القيم الشاذة أو غيرها من التحويلات.

إن الفائدة المتوخاة من التشفير باستعمال التحويل تكمن في تقسيم كمي لبعض المعاملات المحولة، بدقة أكبر دون غيرها من المعاملات الأخرى. وبعبارة أخرى يتم إرفاق عدد كبير من الأرقام الثنائية للمعاملات التي تحتوي على طاقة كبيرة، ويتم تجاهل تلك التي تحوز طاقة ضعيفة. وهذا ما لا يحدث بالنسبة لأنظمة التشفير التي لا تستعمل تحويلًا، بحيث ترفق الأرقام الثنائية بصفة متساوية لجميع المعاملات. من هنا تظهر أهمية التحويل جليا بحيث يستهلك عدد قليل فقط من الأرقام الثنائية.

إن التوصل إلى إنجاز نظام فعال للتشفير يعتمد أساسا على التوفيق بين النوعية الجيدة وإرفاق عدد قليل من الأرقام الثنائية، وهذا سيؤدي حتما إلى ثورة كبيرة في عالم الاتصالات.

إن ما نصبو إليه في المستقبل القريب هو إنجاز نظام إرفاق للأرقام الثنائية الذي يجعل التشفير باستعمال التحويل ذا فعالية قصوى.

أرجو في نهاية هذه الدراسة المتواضعة أن تكون النتائج العلمية التي انتهينا إليها مفيدة للباحثين في هذا الحقل البحثي. كما أتمنى أن تعمل ترجمة هذه الدراسة على ترقية استعمال اللغة العربية من خلال ترجمة المصطلحات الأساسية في ميدان العلاج الآلي للكلام.

ثبت المصطلحات

Communication	الاتصال
Communication téléphonique	الاتصال الهاتفي
Probabilité	احتمال
Probabilité conditionnelle	احتمال شرطي
Probabilité conjointe de transition	احتمال مرفق للانتقال
Probabilité uniforme	احتمال منتظم
Coordonnées	إحداثيات
Statistique	إحصائية
Transmission	الإرسال
Allocation dynamique de bits	الإرفاق الديناميكي للأرقام الثنائية
Résidu	الإشارة المتبقية
Code vectors	أشعة الشفرة
Sons voisés	الأصوات المجهورة
Sons non voisés	الأصوات غير المجهورة
Optimal	أمثل
Sans mémoire	بدون ذاكرة
Dimension	بعد
Structure	بنية
Partition	التجزئة
Echantillonnage	التجزئة إلى عينات
Analyse	التحليل
Analyse numérique	التحليل العددي
Singular Value Decomposition SVD	التحليل إلى القيم الشاذة
Transformation	التحويل

Transformation linéaire	تحويل خطي
Stockage	التخزين
Hierarchique	تدرجي، هرمي
Débit	تدفق
Débit de transmission	تدفق الإرسال
Débit binaire	تدفق ثنائي
Corrélation	الترابط
Synthèse automatique de la parole	التركيب الاصطناعي للكلام
Codage paramétrique	التشفير الوسيط
Codage de la parole	تشفير الكلام
Distorsion moyenne	التشوه المتوسط
Classification automatique	التصنيف الآلي
Orthogonalité	التعامد
Reconnaissance automatique de la parole	التعرف الآلي على الكلام
Pulse Code Modulation (PCM)	التغيير بالنبضات المشفرة
Approximation	تقريب
Quantification	التقسيم الكمي
Quantification vectorielle	التقسيم الكمي الشعاعي
Redondance	التكرار الكلامي
Combinaison linéaire	تنسيق خطي
Fréquence	تواتر
Fréquence d'échantillonnage	تواتر التجزئة
Haute fréquence	تواتر عالي
Basse fréquence	تواتر منخفض
Distribution de la probabilité conjointe	توزيع للاحتمال المرفق
Génération	توليد

Algèbre linéaire	الجبر الخطي
Algèbrique	جبري
Segmental	جزئي
Appareil phonatoire	جهاز النطق
Rapport Signal sur Bruit	حاصل قسمة الإشارة بالنسبة للضجيج
Borne de Shanon	حد "شانون"
Perception	حسية
Erreur de reconstruction	خطا الاسترجاع
Algorithme	خوارزمية
Fonction	دالة
Fonction monotone	دالة رتيبة
Fonction décroissante	دالة متناقصة
Fonction continue	دالة مستمرة
Entropie	درجة التعادل "الحراري"
Entropie conditionnelle	درجة التعادل الشرطية
Bit	رقم ثنائي
Temps discret	زمن متقطع
Amplitude	السعة
Réseaux de points	شبكات النقاط
Variance	شرط التغير
Vecteur	شعاع
Vecteur de base	شعاع قاعدي
Code	شفرة
Compression de la parole	ضغط الكلام
Energie	الطاقة
Longueur	طول

Spectre	الطيف
Matrice colonne	عمود مصفوف
Non corrélé	غير مترابط
Espace euclidien	فضاء إقليدي
Espace des données	فضاء العطايات
Décorrélation	فك الترابط
Dictionnaire	قاموس
Classe	قسم
Block	قطعة
Mean Opinion Score (MOS)	قياس الرأي المتوسط
Mesure subjective	قياس شخصي
Mesure objective	قياس موضوعي
Valeur quadratique moyenne	القيمة الرباعية المتوسطة
Valeur propre	قيمة ضرفة
Valeur discrète	قيمة متقطعة "منفصلة"
Gain	كسب
Quantité	كمية
Quantité d'information	كمية المعلومات
Kilobit	كيلو رقم ثنائي
Corrélé	مترابط
Variabile aléatoire	متغير عشوائي
Ensemble d'apprentissage	مجموعة التمرن
Séquence binaire	مجموعة ثنائية
Ensemble fini	مجموعة منتهية
Contenu informationnel	المحتوى المعلوماتي
Le déterminant	المحدد

Limité	محدود
Transformée	المحولة
Discrete Cosine Transform	محولة DCT
Discrete Fourier Transform	محولة DFT
Filtre de synthèse	مرشح إصدار
Composante	مركبة
Distance Itakura-Saito	مسافة "إيتاكورا-ساييتو"
Distance euclidienne	مسافة إقليدية
Distance cepstrale	مسافة الطيف العكسي
Distance spectrale	مسافة طيفية
Stationnaire	مستقر
Codeur	مشفر
Matrice	مصفوفة
Matrice diagonale	مصفوفة قطرية
Matrice unitaire	مصفوفة موحدة
Matrice unité	مصفوفة وحدة
Equation caractéristique	المعادلة المميزة
Traitement numérique	معالجة رقمية
Coefficients de prédiction linéaire	معاملات التنبؤ الخطي
Coefficients de pondération	معاملات الموازنة
Information mutuelle	المعلومة المتبادلة
Auto-information moyenne	معلومة ذاتية متوسطة
Critère de fidélité	معيار الالتزام
Quantificateur	مقسم كمي
Source gaussienne	منبع "غوسي"
Source prédictible	منبع قابل للتنبؤ

Courbe	منحنى
Région possible	منطقة مسموح بها
Disjoint	منفصل
Onde temporelle	موجة زمنية
Pondéré	موزون
Théorie de la distorsion	نظرية التشوه
Théorie de l'information	نظرية المعلومات
Limite inférieure	النهاية الصغرى

Bibliographie

المراجع

- [1] M. Abbas, "Application de la transformée orthogonale SVD en compression de la parole", thèse de magister, CRSTDLA-ENS, 2002.
- [2] R. Boite, M. Kunt, "Traitement automatique de la parole", Presses Polytechniques Romandes, Lausanne, Suisse, 1987.
- [3] P. Noll, N.S. Jaynt, "Digital Coding Of Waveforms", Pr-Hall, 1984.
- [4] S.K Aoubid, M Boulemden, "Compression du signal de parole par les transformations de Fourier, Cosinus, Hadamard, et en Ondelettes", Proceedings of the IEE'97, International Conference, Vol. Electronic Engineering, 7-9 Dec. 1997, Batna University, Algeria.
- [5] D. Berkani, "Application de la transformée orthogonale SVD en compression de la parole", Séminaire National sur l'Automatique et les Signaux SNAS ' 99, (1999), Annaba University, pp. 132-135.
- [6] L. Knockaert, B. De Backer, and D. De Zutter, "SVD compression, Unitary Transforms, and Computational Complexity", IEEE Transactions on Signal Processing, Vol 47, NO 10, October 1999.
- [7] E. Biglieri, K. Yao, "Some Properties of singular value decomposition and their applications to digital signal processing", Signal Processing Vol. 18 Nov 1989.
- [8] A. Gersho, R. Gray, "Vector quantization and signal compression", Kluwer Academic Publishers, Boston / Dordrecht / London 1992, sixth printing 1997.
- [9] Calliope, "la parole et son traitement automatique", Editeur J.P.Tubach, Masson, Paris, 1989.
- [10] M. Abbas, D. Berkani, "Le codage de la parole par les deux transformées SVD et DCT," Rencontres Jeunes Chercheurs en parole RJC ' 2003, du 23 au 25 septembre 2003, Grenoble, France.
- [11] M. Abbas, D. Berkani, "L'utilisation de la transformation SVD dans la compression des signaux de parole," Séminaire National sur l'Automatique et les Signaux SNAS ' 02, 27 et 28 octobre 2002 Université de Annaba, Annaba, Algérie.

Cahiers du Centre

Série éditée par le Centre de Recherche Scientifique
et Technique pour le Développement de la Langue Arabe

Division Communication parlée et pathologie du langage
Equipe Phonétique et traitement automatique de la parole

Directeur : Rachid BENMALEK

Rédaction du numéro : Mourad ABBAS

Préparation et suivi : Hafnaoui BALI

B.P 255-Alger-Rostomia

16011-Alger

Téléphone : 021941088

Fax : 021793719

<http://www.crstdla.edu.dz>

Cahiers du centre

Série éditée par le

Centre de Recherche Scientifique et Technique
pour le Développement de la Langue Arabe

Notions sur le codage de la parole

Application de la transformation SVD

Mourad Abbas

N° 6

2007

Dépôt Légal : 2007-3383

ISSN : 1112-735x

ISBN : 978-9961-9660-5-1

Résumé

De nos jours, le codage de la parole est sans nul doute, une opération cruciale dans un domaine stratégique comme les communications. Nous présentons dans cet article des notions importantes relatives à ce thème et exposons la méthode que nous avons choisie pour réaliser la compression du signal vocal.

En premier lieu nous présentons les notions de base issues de la théorie de l'information, ainsi que les opérations élémentaires qui précèdent le codage, et qui sont généralement requises pour le traitement du signal vocal, comme l'échantillonnage et la quantification. La méthode que nous avons adoptée est l'hybridation entre deux types de codage que sont le codage par transformée en utilisant la transformation orthogonale SVD (Singular Value Decomposition) et le codage paramétrique. Ce dernier est lié, dans notre cas, aux coefficients de prédiction linéaire qui sont des paramètres largement utilisés dans le traitement de la parole. Dans les systèmes basés sur le codage par transformée, chaque bloc d'échantillons du signal vocal est transformé en un ensemble de paramètres qui seront quantifiés indépendamment, avant leur transmission. La quantification vectorielle des coefficients résultants de la transformation est assurée par le quantificateur LBG.

A la réception, le signal est restitué par le biais d'un filtre de synthèse excité par le résidu. La qualité du signal obtenu est jugée très satisfaisante.

Mots clés : codage de la parole, Transformation SVD, Quantification vectorielle.

Sommaire

1. Introduction.....	9
2. Notions sur la théorie de la distorsion.....	10
2.1. Entropie d'une source.....	10
2.2. La fonction de distorsion.....	11
3. Représentation numérique du signal.....	15
3.1. La numérisation : Définition simplifiée.....	16
4. Mesure de la qualité d'un codeur.....	19
4.1. Mesure objective.....	19
4.2. Mesure subjective.....	19
5. Notion de transformation.....	19
5.1. Interprétation.....	24
6. Codage par transformation.....	24
6.1. Allocation de bits.....	25
7. la transformation SVD.....	26
7.1. Valeurs et vecteurs propres.....	27
7.2. Calcul de la SVD d'une matrice.....	28
7.3. Interprétation de la transformation SVD.....	29
8. Le codage par SVD.....	30
9. Quantification vectorielle.....	33
9.1. Définition.....	33
9.2. Quantification vectorielle statistique.....	34
9.3. Quantification vectorielle algébrique.....	35
9.4. Algorithme LBG.....	36
9.5. Performances du quantificateur.....	37
10. Expérimentation et résultats.....	43
11. Conclusion.....	49
Bibliographie.....	50

1. Introduction

La parole a toujours été le principal moyen de communication entre les humains. Aujourd'hui les recherches sont actives pour arriver à réaliser ce que l'on appelle dialogue Homme-Machine. Ce processus exige la maîtrise de trois axes principaux qui sont: la reconnaissance, la synthèse, et le codage de la parole.

Le codage d'un signal n'est autre qu'une mise en forme ou une représentation appropriée en vue de son stockage ou sa transmission.

Les sons de la parole nécessitent des capacités de stockage importantes et des débits de transmission élevés. Leur compression s'avère alors nécessaire pour qu'ils occupent moins d'espace et pour qu'ils aient un débit réduit. Pour ce faire, on réduit la quantité d'informations contenue dans le signal en exploitant les redondances intrinsèques à ces informations, ou en adaptant le contenu informationnel aux capacités limitées du système utilisateur (l'oreille humaine). Pour une meilleure compréhension, nous présentons dans la section 2 la notion de la théorie d'information et celle de la théorie de distorsion qui en découle. Nous montrons ensuite les étapes élémentaires que nous devons appliquer au signal vocal avant de le coder.

La notion de transformation et d'orthogonalité seront présentées et illustrées par un exemple dans la section 3. Nous introduisons également la définition de la transformée SVD : Singular Value Decomposition, ses propriétés et ses caractéristiques, qui justifient son efficacité dans le codage de la parole.

Nous présentons ensuite la façon dont un système de codage par transformation se réalise, et particulièrement celui qui utilise la SVD. La quantification vectorielle est une étape élémentaire dans un tel système, pour cela nous avons étudié les

performances de notre quantificateur en introduisant à son entrée un signal gaussien que nous avons généré au préalable. La dernière étape qui est celle de l'évaluation des signaux synthétisés est réalisée en utilisant deux mesures, l'une est objective tandis que l'autre est dite subjective.

2. Notions sur la théorie de la distorsion

2.1. Entropie d'une source

L'entropie d'une source $X=(x(0), x(1), \dots, x(N-1))$, à temps discret, stationnaire, ergodique et sans mémoire, notée par $H(x)$ est définie comme étant l'auto - information moyenne de cette source [1], et est exprimée par la relation:

$$H(X) = E[I(X)] \quad (2.1)$$

$I(X)$ est l'auto-information de la source, exprimée en bits par l'équation (2.2):

$$I(X) = -\log_2 P(X) \quad (2.2)$$

Chaque échantillon prend ses valeurs dans l'ensemble $\{a_0, a_1, \dots, a_{N-1}\}$, et est associé à une probabilité définie par :

$$p(x(i)) = a(i) = p_X(a_i).$$

Les probabilités des différents éléments de X sont définies par : $p(a_0), p(a_1), \dots, p(a_{N-1})$. La probabilité $P(X)$ permet donc de définir l'auto-information $I(x)$ dont nous avons parlé. De (2.1) et (2.2) on a l'équation :

$$H(X) = -E[\log_2 P(X)] = -\sum_{i=0}^{N-1} P_X(a_i) \log_2 P_X(a_i) \quad (2.3)$$

sachant que :

$$0 \leq P_X(a_i) \leq 1 \quad \sum_{i=0}^{N-1} P_X(a_i) = 1 \quad (2.4)$$

L'entropie d'une source est toujours positive et vérifie:

$$0 < H(X) < \log_2 N$$

Lorsque $H(x) = \log_2 N$ on obtient une probabilité uniforme $P_X = 1/N$, et la source est totalement non prédictible. Lorsque $H(x)$ est nulle, la source est totalement prédictible.

Dans le cas de deux variables aléatoires X et Y associées aux alphabets de valeurs :

$\{a_0, a_1, \dots, a_{N-1}\}$ et $\{b_0, b_1, \dots, b_{M-1}\}$ avec une distribution de probabilité conjointe $p_{XY}(a_i, b_j)$, l'entropie conditionnelle, déduite de la notion de probabilité conditionnelle, exprime l'information d'une source connaissant une autre source. Elle est donnée par la relation suivante:

$$\begin{aligned} H(X, Y) &= -E[\log_2 P_{X/Y}(X/Y)] = \\ &= -\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_{XY}(a_n, b_m) \log_2 P_{X/Y}(a_n / b_m) \end{aligned} \quad (2.5)$$

2.2. La fonction de distorsion

Dans un quantificateur, l'approximation de la séquence d'entrée, en l'occurrence le vecteur X , par un autre vecteur Y , introduit une distorsion. Cette dernière est mesurée par la relation suivante qui exprime le critère de fidélité :

$$\rho_k(x, y) = \frac{1}{k} \sum_{i=0}^{k-1} \rho_k(x(i), y(i)) \quad (2.6)$$

La quantité ρ est donc une mesure de distorsion entre les composantes des vecteurs X et Y .

La distorsion moyenne pour la fonction de probabilité conjointe de transition est donnée par l'expression :

$$d(P_{Y/X}) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_X(a_n) P_{Y/X}(b_m/a_n) \rho(a_n, b_m) \quad (2.7)$$

L'information mutuelle qui est considérée comme étant la quantité d'information apprise sur Y en observant les valeurs de X est donnée par la relation (2.8):

$$I(P_{Y/X}) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_X P_{Y/X}(b_m/a_n) \log_2 \frac{P_{Y/X}(b_m/a_n)}{P_Y(b_m)} \quad (2.8)$$

$$\text{avec } P_{Y/X} = \sum_{n=0}^{N-1} P_X(a_n) P_{Y/X}(b_m/a_n) \quad (2.9)$$

La fonction de distorsion $R(D)$ dans la relation (2.10), donne un débit minimal R pour une distorsion donnée D , ou inversement pour un débit donné, elle fournit la plus faible distorsion possible D .

$$R(D) = \min_{P_{Y/X} \in \mathcal{P}_D} I(P_{Y/X}) \quad (2.10)$$

$$\text{avec } \mathcal{P}_D = \left\{ P_{Y/X} : d(P_{Y/X}) \leq D \right\}$$

Un code C de taille N et de longueur de block k est un ensemble de N séquences de dimension k chacune. Ces dernières sont des séquences binaires appelées mots de code de longueur $\log_2 N$.

Si nous voulons coder un vecteur de k échantillons nous lui affectons un débit R par échantillon:

$$R = \frac{\log_2 N}{k} \quad (2.11)$$

La fonction de distorsion $R(D)$ est une fonction continue, monotone et décroissante. Elle représente la limite théorique des performances des systèmes de codage. L'équation (2.12) l'exprime comme suit:

$$R(D) = \lim_{k \rightarrow \infty} \frac{1}{k} \log_2 N \quad (2.12)$$

En pratique, il est possible d'atteindre une performance assez proche de cette limite.

La figure (2.1) montre la courbe représentative de la fonction de distorsion dans le cas d'une source discrète. Les équations citées ci-dessus sont toutes relatives à cette dernière.

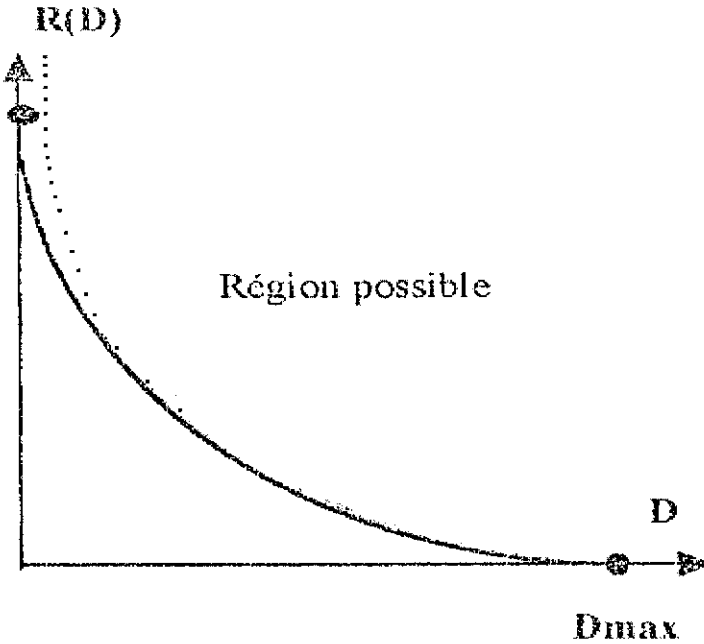


Fig 2.1. La fonction de distorsion $R(D)$ pour le cas discret (en trait continu) et celle en cas d'une source continue (en pointillé).

La figure (2.1) montre qu'il existe une région possible qui est délimitée par la limite inférieure de la fonction $R(D)$ appelée borne de Shannon, dont il est interdit de passer en dessous. Nous remarquons aussi que pour une distorsion nulle, $R=H$. Par contre pour une valeur maximale D_{max} nous avons $R=0$. Pour le cas d'une source continue, comme il est indiqué dans la figure (2.1), la courbe $R(D)$ tend vers l'infini pour une distorsion faible.

3. Représentation numérique du signal

Pour représenter numériquement un signal, on doit tout d'abord lui appliquer un échantillonnage à une fréquence F_s , en vertu du théorème de Shannon. Chaque échantillon sera quantifié selon un nombre fini de valeurs discrètes. Le signal, avec sa nouvelle représentation, sera ainsi codé avec la méthode adéquate.

Le spectre du signal vocal peut s'étendre jusqu'à 12 kHz [2]; en se basant sur le critère de Shannon, la fréquence d'échantillonnage appropriée est 24kHz.

Cependant dans le domaine des communications on limite le spectre afin de diminuer le coût de l'opération du traitement numérique. Dans le cas de la téléphonie, la qualité est jugée suffisante pour un spectre ne dépassant pas 3400 Hz, et la fréquence d'échantillonnage adoptée est de 8 kHz.

Dans le domaine de l'analyse, la reconnaissance et la synthèse du signal vocal, cette fréquence varie entre 6 kHz et 16 kHz.

Pour la téléphonie, les échantillons quantifiés sont codés par des mots de 8 bits. Le débit binaire correspondant vaut 64 kb/s étant donné que le signal est échantillonné à 8kHz. Ceci est relatif à la méthode PCM (Pulse Code Modulation).

Les algorithmes de codage de la parole qui utilisent une modélisation du système phonatoire (gorge, cordes vocales, ...etc) permettent d'obtenir des débits avoisinant 2 kilobits par seconde, toutefois la qualité est médiocre car la distorsion est grande.

Le but des diverses techniques de codage apparues à ce jour est d'aboutir à un système réalisant un compromis entre la performance et la qualité, le débit binaire, le coût (prix), et la complexité du traitement.

3.1. La numérisation : Définition simplifiée

A l'origine, le signal électrique acquis est analogique. Pour mieux comprendre ce que veut dire la numérisation, nous avons préféré donner quelques définitions simples illustrées par des schémas.

Définition 1 : Un signal analogique a des valeurs d'amplitude et de temps continues, il est illustré par la figure (3.1).

Définition 2 : Quantifier un signal analogique c'est simplement quantifier son amplitude, le temps reste continu. Voir figure (3.2).

Définition 3 : Un signal à amplitude continue et à temps discret est appelé signal échantillonné, il est représenté dans la figure (3.3).

Définition 4 : Le signal schématisé dans la figure (3.4) à amplitude et temps discrets est appelé signal numérique.

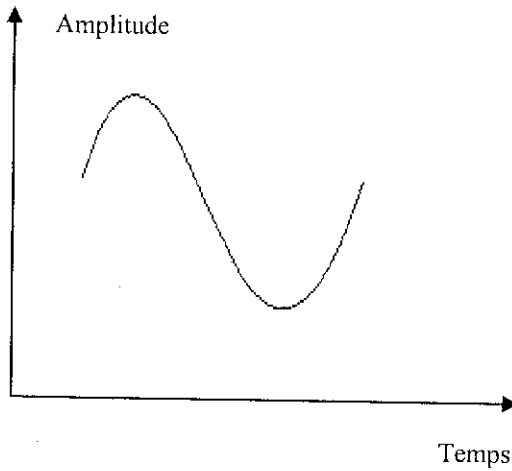


Fig 3.1.Exemple d'un signal analogique

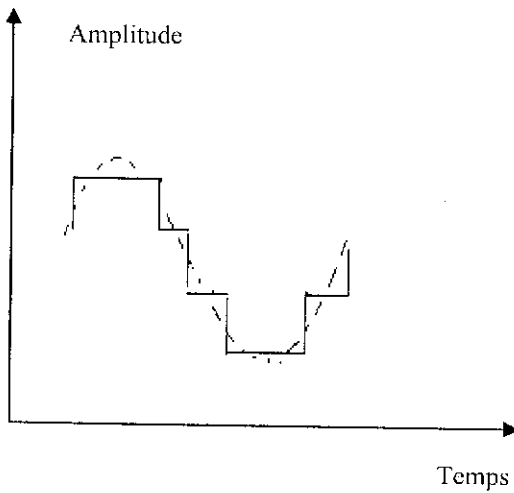


Fig 3.2. Discretisation de l'amplitude du signal

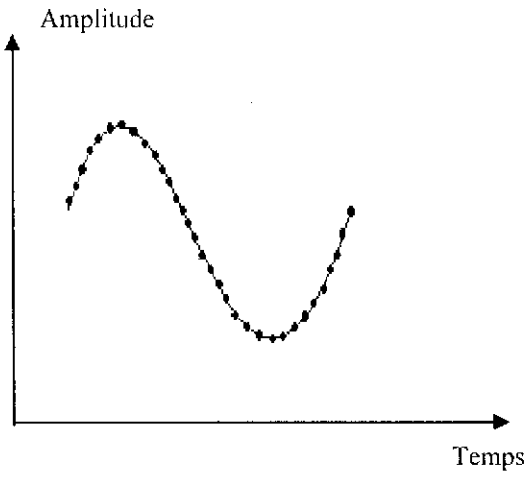


Fig 3.3. Discrétisation du temps

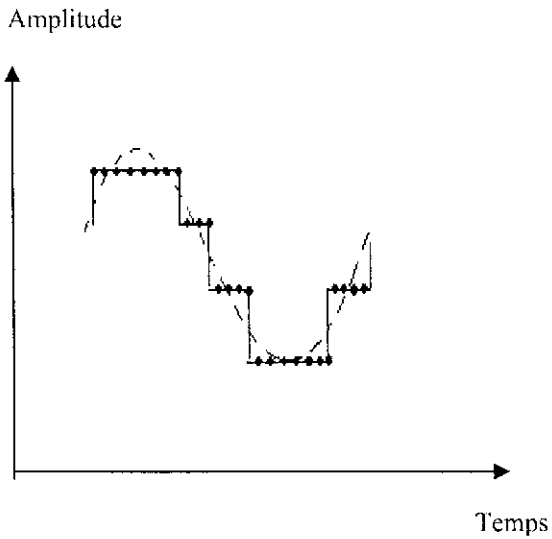


Fig 3.4. Discrétisation du temps et de l'amplitude.

4. Mesure de la qualité d'un codeur

4.1. Mesure objective

La mesure objective standard de la qualité d'un codeur est le rapport de la variance du signal sur la variance de l'erreur, appelé rapport signal sur bruit, (SNR:Signal Noise Ratio) mesuré en décibels

$$SNR = 10 \log \frac{\sigma_s}{\sigma_e} \quad (4.1)$$

Toutefois cette notion s'est révélée expérimentalement inadéquate pour mesurer la qualité perceptuelle d'un codeur de qualité, c'est pour cela que le SNR segmental a été introduit. Ce dernier n'est rien d'autre que la moyenne des SNR calculés pour chaque trame du signal. Il existe d'autres critères d'évaluation comme la distance d'Itakura pondérée.

4.2. Mesure subjective

L'écoute du signal est une mesure subjective, elle s'ajoute aux mesures objectives qui ne peuvent pas donner toujours l'information correcte de la qualité du codeur perçue par l'utilisateur, du fait qu'elles ne tiennent pas compte des propriétés perceptuelles de l'oreille. Cette mesure est plus détaillée dans la section 10.

5. Notion de transformation:

La notion de transformation peut être expliquée et illustrée par l'exemple suivant [3]:

Soit la séquence du signal $\{x(n)\}$ présentée dans la figure (5.1) suivante:

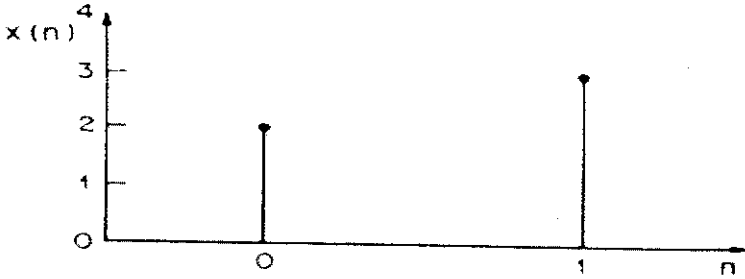


Fig 5.1. La séquence $x(n)$ [3]

Cette transformation est d'ordre 2:

$$x(0)=2 ; x(1)=3 \quad (5.1)$$

On peut écrire la séquence $\{x(n)\}$ sous une forme représentative et approximative: $\{5/2, 5/2\}$;

$$\{5/2, 5/2\} = \frac{5}{\sqrt{2}} \left\{ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\} = \frac{5}{\sqrt{2}} \{b_{00}, b_{01}\} \quad (5.2)$$

Le signal $\{x(n)\}$ peut être exactement décrit en ajoutant la séquence $\{-0.5, +0.5\}$ à l'équation (5.2):

$$\{-0.5, +0.5\} = \frac{-1}{\sqrt{2}} \left\{ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\} = \frac{-1}{\sqrt{2}} \{b_{10}, b_{11}\} \quad (5.3)$$

Les vecteurs b_0 et b_1 qui sont définis dans (5.2) et (5.3) et montrés respectivement dans les figures (5.2) et (5.3) ci-dessous, peuvent être considérés comme des vecteurs de base, représentant la séquence $\{x(n)\}$ comme une combinaison linéaire pondérée par des coefficients de transformation :

$$T(0) = \frac{5}{\sqrt{2}} = \frac{1}{\sqrt{2}}(x(0) + x(1)) \quad (5.4)$$

$$T(1) = \frac{-1}{\sqrt{2}} = \frac{1}{\sqrt{2}}(x(0) - x(1)) \quad (5.5)$$

Les coefficients transformés $T(0)$ et $T(1)$ représentent respectivement, la somme et la différence des deux échantillons de l'entrée $x(n)$, en d'autres termes, l'énergie des composantes basse fréquence et des composantes haute fréquence dans la séquence $x(n)$.

On déduit que l'isolation des composantes basse fréquence et celles des hautes fréquences est la motivation de la représentation par transformation [3].

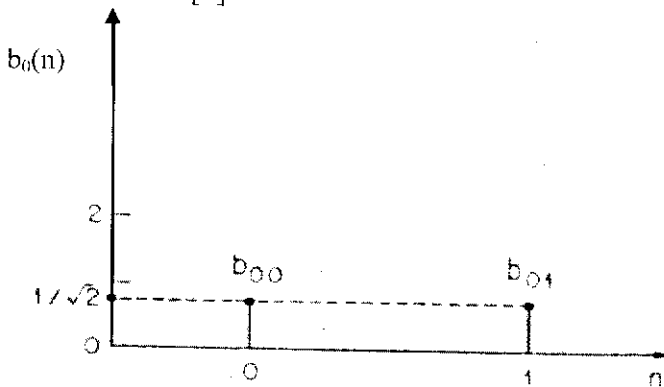


Fig 5.2. Le vecteur de base b_0 [3]

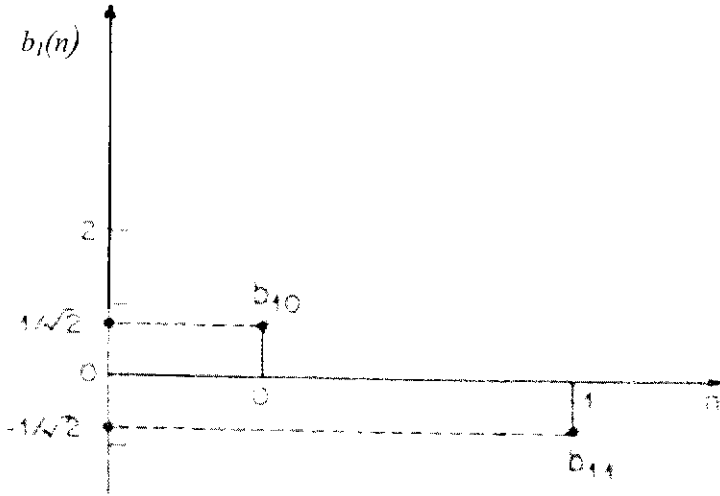


Fig 5.3. Le vecteur de base : b_j [3]

Nous pouvons reconstruire les valeurs $x(0)$ et $x(1)$ par l'opération inverse suivante :

$$x(0) = \frac{1}{\sqrt{2}}(T(0) + T(1)) \quad (5.6)$$

$$x(1) = \frac{1}{\sqrt{2}}(T(0) - T(1)) \quad (5.7)$$

L'opération de transformation fournit en général des coefficients non corrélés avec des variances différentes. On peut facilement remarquer que pour une séquence aléatoire stationnaire $x(n)$ transformée par les équations (5.4) et (5.5) on a :

$$E(T(0)T(1)) = E[x(0)^2] - E[x(1)^2] = 0 \quad (5.8)$$

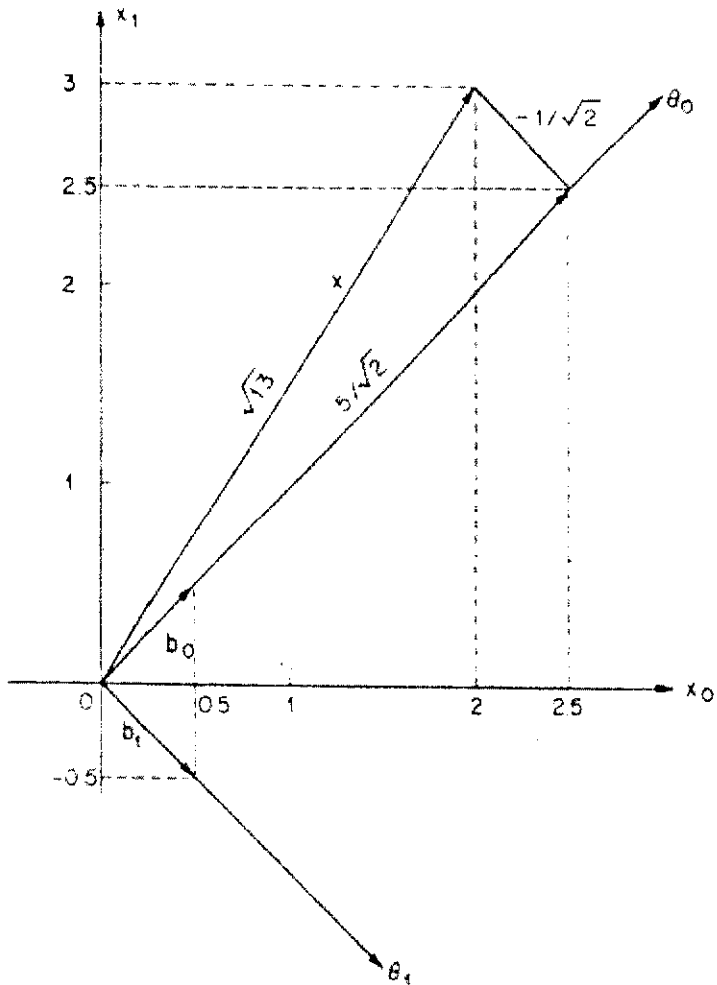


Fig 5.4. Interprétation géométrique de la représentation de la transformation de l'entrée (x_0, x_1) [3]

5.1. Interprétation:

La décorrélation obtenue par l'opération de transformation est illustrée géométriquement dans la figure (5.4) afin de bien assimiler cette notion.

Considérons l'entrée $x(n)$ comme un point dans le domaine à deux dimensions (x_0, x_1) , on remarque que les vecteurs b_0 et b_1 sont orthogonaux, ils constituent un nouveau domaine (T_0, T_1) ou le point (x_0, x_1) peut être représenté par les coordonnées $(T(0), T(1))$.

6. Codage par transformation

La technique de compression par transformée consiste à traiter le signal d'entrée numérique au moyen des transformations orthogonales et à coder la sortie transformée pour réduire la quantité d'information [4].

Nous citons à titre d'exemple quelques transformations qui peuvent être utilisées, la SVD (Singular Value Decomposition), la DCT (Discrete Cosine Transform) qui sont sous-optimales et la DFT (Discrete Fourier Transform) [5].

L'idée du codage par transformée consiste à appliquer une transformation linéaire à une séquence de signal corrélée pour obtenir un ensemble de coefficients non corrélés.

Une partie de ces coefficients, celle qui véhicule la plus grande partie de l'information, sera quantifiée puis codée, et il est connu que la plupart des transformations procèdent à l'élimination des coefficients à énergie minimale, il est préférable donc d'utiliser une transformation qui comprime le maximum d'énergie dans un nombre minimum de coefficients, pour pouvoir choisir les coefficients à transmettre.

Pour ce faire, plusieurs méthodes peuvent être appliquées, nous pouvons citer la méthode dite du seuillage. Cette dernière prend en considération les coefficients transformés en déterminant un seuil sur leurs amplitudes et éliminant les coefficients dont l'amplitude n'atteint pas ce seuil [4].

Une autre méthode est le calcul de la variance de chaque composante de transformation et d'appliquer l'algorithme d'allocation dynamique de bits afin de quantifier avec plus de résolution certaines composantes [5, 6].

L'efficacité d'un système de codage par transformation dépend du type de transformation et de l'allocation de bits au cours du codage [6].

Pour réaliser la compression, il est important de suivre la transformation avec un dispositif d'allocation de bits afin de quantifier et de coder chaque coefficient de séquence et ainsi effectuer une meilleure compression. La figure (6.1) illustre le processus :

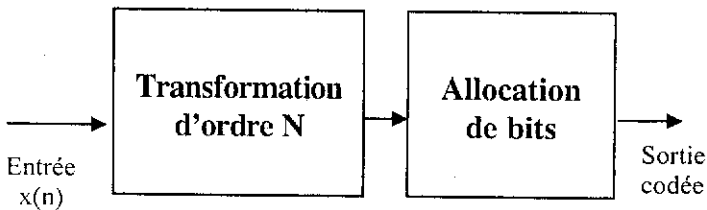


Fig 6.1. Compression par transformée

6.1. Allocation de bits

L'allocation de bits est la partie la plus importante dans le codage par transformation. Elle consiste à minimiser la valeur quadratique moyenne de l'erreur de reconstruction et permet de quantifier avec plus de résolution certaines composantes,

c'est ainsi que l'on obtient un maximum de gain par rapport à la méthode de codage simple.

Les valeurs des variances des composantes de transformation sont différentes. En effet, elles ont des débits R_k différents. Le débit moyen R est constant, il s'exprime par la relation (6.1) :

$$R = \frac{1}{N} \sum_{k=0}^{N-1} R_k = \text{constante} \quad (6.1)$$

La représentation de l'allocation de bits optimale s'exprime sous la forme :

$$R_k = R + \frac{1}{2} \log_2 \frac{\sigma_k^2}{\left[\prod_{j=0}^{N-1} \sigma_j^2 \right]^{\frac{1}{N}}} \quad (6.2)$$

Comme le montre la relation précédente, l'allocation de bits optimale dépend de la distribution des variances des composantes de transformation. Le plan de l'allocation de bits est détaillé dans [5].

7. La transformation SVD

La décomposition en valeurs singulières SVD est un des outils de base les plus importants en algèbre linéaire et trouve de plus en plus d'applications en traitement du signal.

Dans ce qui suit nous allons donner quelques propriétés de la SVD dans le cas des matrices.

La SVD est la généralisation de la représentation classique par les valeurs et les vecteurs propres d'une matrice carrée [7]. Elle est utilisée pour résoudre certains problèmes d'analyse numérique.

7.1. Valeurs et vecteurs propres

Soit un système de n équations linéaires du type :

$$AX = \lambda X \quad \text{ou} \quad a_{ij}x^j = \lambda x^i, \quad (7.1)$$

avec λ un scalaire.

A est une matrice carrée et X une matrice colonne. On peut écrire cette équation sous la forme (7.2) :

$$(A - \lambda U)X = 0 \quad \text{ou} \quad (a_{ij} - \lambda \delta_{ij})x^j = 0 \quad (7.2)$$

Avec U la matrice unité d'ordre n .

La solution $X=0$ ne présente aucun intérêt. Pour que le système admette d'autres solutions non nulles, il faut et il suffit que le déterminant soit nul :

$$\|A - \lambda U\| = 0 \quad \text{ou} \quad \|a_{ij} - \lambda \delta_{ij}\| = 0 \quad (7.3)$$

Ou encore :

$$\begin{vmatrix} (a_{11} - \lambda) & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & (a_{22} - \lambda) & & \dots & a_{2n} \\ a_{31} & & (a_{33} - \lambda) & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & (a_{nn} - \lambda) \end{vmatrix} = 0 \quad (7.4)$$

En développant ce déterminant on a un polynôme de degré n qui s'écrit:

$$P(\lambda) = (-1)^n \cdot [\lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \dots + a_{n-1} \lambda + a_n] \quad (7.5)$$

Il est appelé le polynôme caractéristique, tandis que l'équation $P(\lambda) = 0$ est l'équation caractéristique de A .

Les racines λ_i de ce polynôme, au nombre de n , sont des solutions non nulles de (7.1), elles sont les valeurs propres de la matrice A .

A chaque valeur propre λ_i correspond une solution non nulle de (7.1), X_i , appelée vecteur propre de A associé à λ_i .

7.2. Calcul de la SVD d'une matrice

Toute matrice complexe donnée A de taille $M \times N$, de rang $M \geq N$, admet la représentation suivante:

$$A = USV^T; \quad (7.6)$$

V^T est la transposée conjuguée de V

S est une matrice diagonale de taille $N \times N$

$$S = \text{diag}(s_0, s_1, \dots, s_{N-1}) \quad (7.7)$$

Les éléments diagonaux de cette matrice sont réels positifs et sont mis dans l'ordre décroissant :

$$S_0 \geq S_1 \geq \dots \geq S_{N-1} > 0 \quad (7.8)$$

U et V sont deux matrices unitaires constituées de taille $N \times N$:

$$U = (U_0, U_1, \dots, U_{N-1}) \quad (7.9)$$

$$V = (V_0, V_1, \dots, V_{N-1}) \quad (7.10)$$

$$U U^T = U^T U = I_N; \quad U^T = U^{-1} \quad (7.11)$$

$$V V^T = V^T V = I_N; \quad V^T = V^{-1} \quad (7.12)$$

I_N est la matrice identité de taille $N \times N$

La décomposition en valeurs singulières de la matrice A peut être écrite sous la forme suivante:

$$A = \sum_{i=0}^{N-1} S_i U_i V_i^T \quad (7.13)$$

Les vecteurs propres de la matrice $H^T H$ constituent les colonnes de la matrice unitaire V , et les vecteurs propres de la matrice $H H^T$ forment les colonnes de la matrice unitaire U .

Les deux matrices U et V ont les mêmes valeurs propres qui sont égales au carré des valeurs S_i de la matrice S .

7.3. Interprétation de la transformation SVD

Nous avons vu précédemment que la SVD d'une matrice A de taille $N \times N$ est définie comme suit :

$$A = \sum_{i=0}^{N-1} S_i U_i V_i^T$$

Cette équation peut être développée sous la forme (7.14) :

$$\begin{aligned} [a_0, a_1, \dots, a_{N-1}] &= S_0 U_0 V_0^T + \dots + S_{N-1} U_{N-1} V_{N-1}^T \\ &= S_0 [V_{00} U_0, \dots, V_{0N-1} U_0] + S_1 [V_{10} U_1, \dots, V_{1N-1} U_1] + \\ &\quad \dots + S_{N-1} [V_{N-1,0} U_{N-1}, \dots, V_{N-1,N-1} U_{N-1}] \end{aligned} \quad (7.14)$$

En regardant chaque colonne de la matrice A , on remarque facilement que :

$$\alpha_i = \sum_{j=0}^{N-1} S_j V_{ji} U_j \quad i = 1, 2, \dots, N-1. \quad (7.15)$$

Fournir la SVD d'une matrice A revient donc à écrire chaque colonne de cette matrice sous la forme d'une somme pondérée des vecteurs de base U_j , ($j=0, \dots, N-1$). Pour la colonne a_i , les coefficients de pondération sont le produit des valeurs singulières par les composantes de la $i^{\text{ème}}$ colonne du vecteur V_j .

8. Le codage par SVD

La transformation SVD joue un rôle considérable dans le traitement du signal. Elle a la puissance d'extraire l'espace des données noyé dans un bruit [6].

En ce qui concerne la parole, elle permet la décorrélation du signal vocal en séparant les sons voisés des sons non voisés (les signaux à haute énergie de ceux à énergie basse).

Le système de codage que nous avons appliqué, consiste à calculer la SVD d'une matrice A qui représente les coefficients de prédiction $a(i)$ et d'extraire la matrice V , qui est la matrice de

transformation dans notre cas. Il est nécessaire que la matrice de transformation soit orthogonale et en général unitaire pour le codage par transformation [5]. Nous appliquons ensuite la quantification vectorielle sur les coefficients transformés (les éléments de la matrice V). Au niveau du récepteur on reconstruit les coefficients de prédiction $\hat{a}(i)$ à partir de ces coefficients transformés transmis. Ils sont donnés par la relation suivante :

$$A = USV^T$$

A est la matrice contenant les coefficients $\hat{a}(i)$ restitués. Le schéma (8.1) représente le système de codage par SVD.

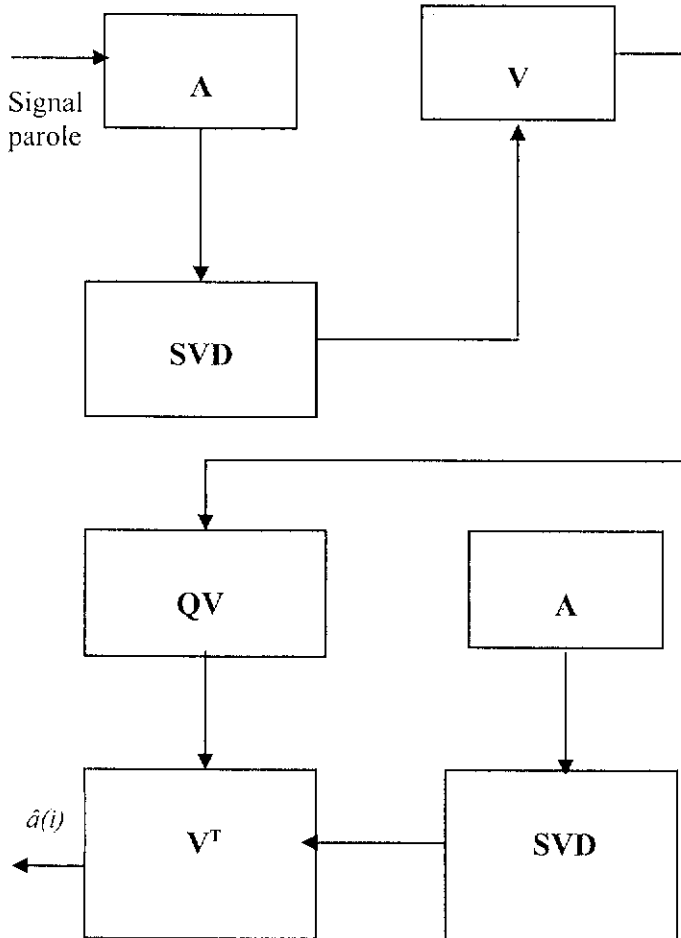


Fig 8.1. Système de codage par SVD

9. Quantification vectorielle :

Dans cette partie nous allons définir la quantification vectorielle VQ (Vector Quantization). Nous présentons aussi l'une des méthodes utilisées pour réaliser cette opération; en l'occurrence le quantificateur LBG, tout en exposant ses performances.

La VQ est une technique utilisée dans la compression des signaux. Dans les toutes dernières années, elle est devenue un outil important dans la compression de la parole et de l'image, ainsi que dans la reconnaissance de la parole. Son importance dans ces domaines croit de jour en jour.

La quantification vectorielle est la conséquence du passage de la quantification scalaire à celle d'un vecteur [8].

9.1. Définition

Un quantificateur vectoriel Q de dimension k et de taille N est une application d'un vecteur appartenant à un espace euclidien à k dimensions, E_k , vers un ensemble fini C contenant N vecteurs de sortie appelés code vectors ou code words.

$$Q : E_k \rightarrow C$$

avec $C = \{y_1, y_2, y_3, \dots, y_N\}$ et $y_i \in E_k$ pour tout $i \in I = \{1, 2, \dots, N\}$.

La quantification vectorielle peut être statistique ou algébrique.

La VQ permet de représenter un vecteur x à k composantes par un autre vecteur y qui appartient à un ensemble fini de M vecteurs représentant le dictionnaire. La substitution de x par y entraîne une distorsion ou distance notée $d(x, y)$.

Il existe diverses méthodes de calcul de distorsion, nous citons la distance spectrale, la distance cepstrale et la distance de Itakura-Saito.

La distance la plus utilisée est la distance euclidienne mentionnée ci-dessous par l'équation (9.1):

$$d_2(x, y) = \frac{1}{k} \sum_{i=1}^k (x_i - y_i)^2 \quad (9.1)$$

la quantification vectorielle a pour but de minimiser la distorsion moyenne pour un nombre de classes donné, ou encore de réduire le nombre de classes en fixant la distorsion moyenne.

Pour cela de nombreuses méthodes sont utilisées dont l'Algorithme de Lloyd généralisée (k-means) et l'algorithme par éclatements binaires LBG qui sera présenté ci-après.

9.2. Quantification vectorielle statistique

Il existe deux types de classification automatique, les unes hiérarchiques, les autres non hiérarchiques. Les méthodes hiérarchiques sont en général peu utilisées en traitement de la parole.

La quantification vectorielle statistique est une méthode non hiérarchique qui aboutit à une seule partition des données en classes : l'ensemble d'apprentissage est alors décomposé en sous-ensembles disjoints.

Chacun de ces sous-ensembles est en général caractérisé par son centre de gravité appelé centroïde, et sert au cours des calculs à trouver la partition finale.

Dans la figure (9.1) on remarque qu'une classe est définie par l'ensemble des points les plus proches de son centre de gravité. Les centroïdes et l'ensemble de données sont représentés respectivement par les étoiles et les petits cercles.

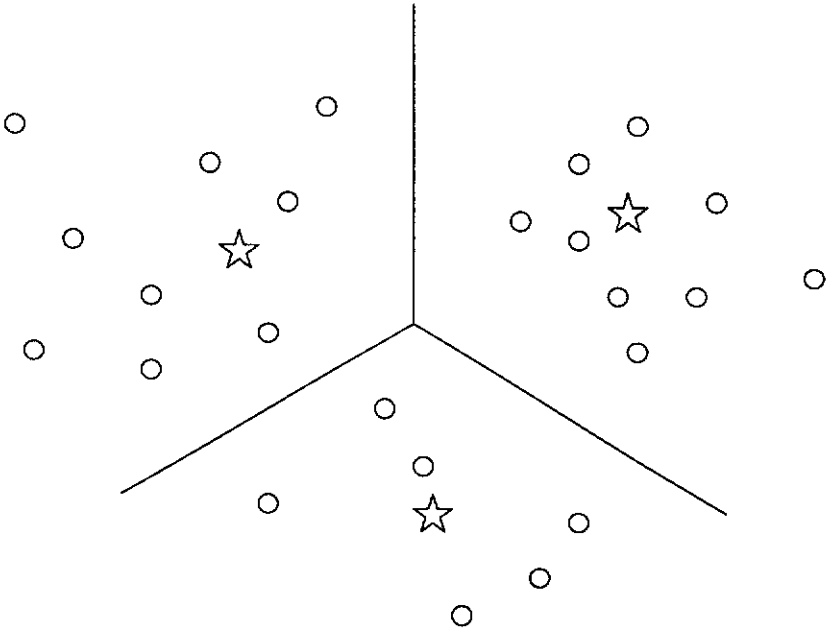


Fig 9.1. Représentation des classes

9.3. Quantification vectorielle algébrique

Le principe de cette méthode est d'utiliser un dictionnaire qui possède une structure mathématique forte. Cette approche tire profit des structures algébriques régulières telles que les codes, les réseaux de points et les treillis pour concevoir des quantificateurs vectoriels performants, peu complexes et virtuellement sans besoin de stockage.

9.4 Algorithme LBG

Considérons un ensemble de L vecteurs x qui doit être réparti judicieusement en un nombre de M classes imposé a priori [9].

L'algorithme consiste à :

- Choisir M centroïdes d'une façon aléatoire dans l'espace E_k .
- Affecter chaque vecteur x à sa classe en calculant les distances entre ce vecteur et tous les centroïdes. Le vecteur x est alors associé au centroïde le plus proche.
- Recalculer la position de chaque centroïde y_i , et cela en considérant le centre de gravité de chaque nouvelle classe comme centroïde, afin de minimiser la distorsion D_i avec :

$$D_i = \sum_j d(x_j^{(i)}, y_i) \quad (9.2)$$

avec $d(x_j^{(i)}, y_i)$ représente la mesure de distorsion entre $x_j^{(i)}$ et y_i .

$x_j^{(i)}$ représente les vecteurs qui appartiennent à la classe i et y_i est le vecteur prototype ou centroïde.

- Calculer la distorsion totale D :

$$D = \sum_{i=1}^M D_i \quad (9.3)$$

- Itérer jusqu'à ce que la distorsion D varie de moins de ε d'une itération à l'autre. ε représente un seuil qu'on détermine selon la précision voulue.

Dans cette méthode, les centroïdes initiaux sont choisis d'une façon empirique, il arrive qu'un choix s'avère meilleur qu'un autre car il mène à une distorsion plus faible. il est donc recommandé d'essayer plusieurs choix initiaux différents.

9.5. Performances du quantificateur

Pour montrer les performances du quantificateur, nous avons généré un signal gaussien, sur lequel nous avons appliqué la quantification vectorielle pour différentes tailles du dictionnaire. Chaque vecteur contient dix valeurs du signal gaussien. Les figures suivantes présentent chacune le rapport signal sur bruit (RSB) obtenu pour des tailles de dictionnaire différentes.

RSB en dB

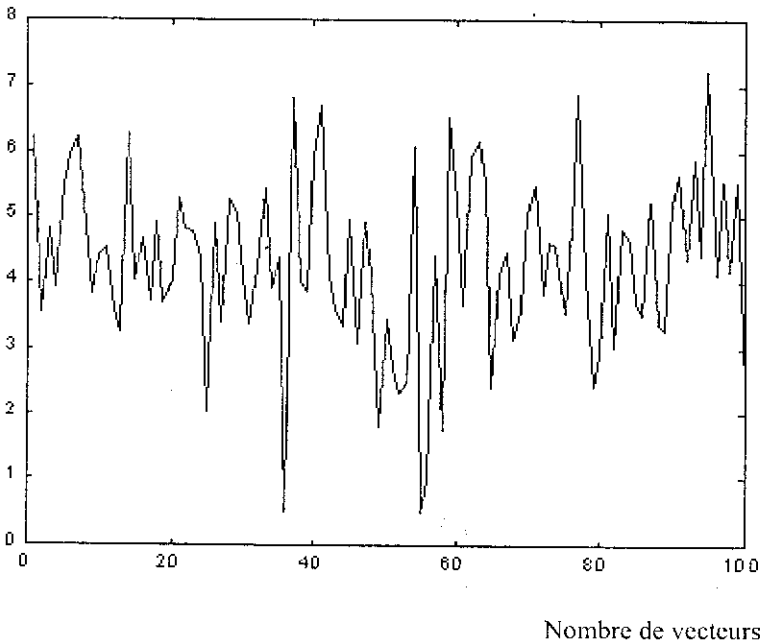


Fig 9.2. Rapport signal sur bruit pour un dictionnaire de taille 512 points

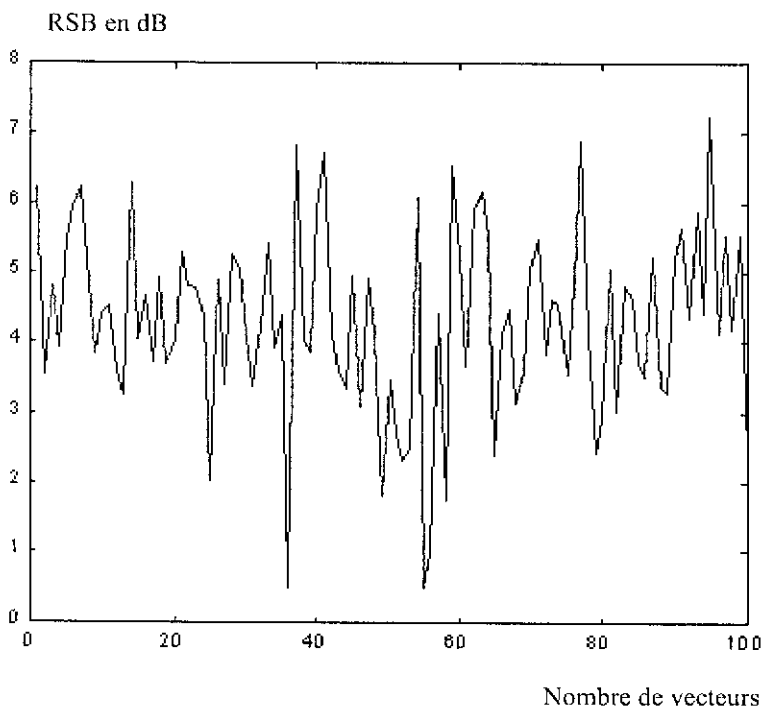


Fig 9.3. Rapport signal sur bruit pour un dictionnaire de taille 256 points.

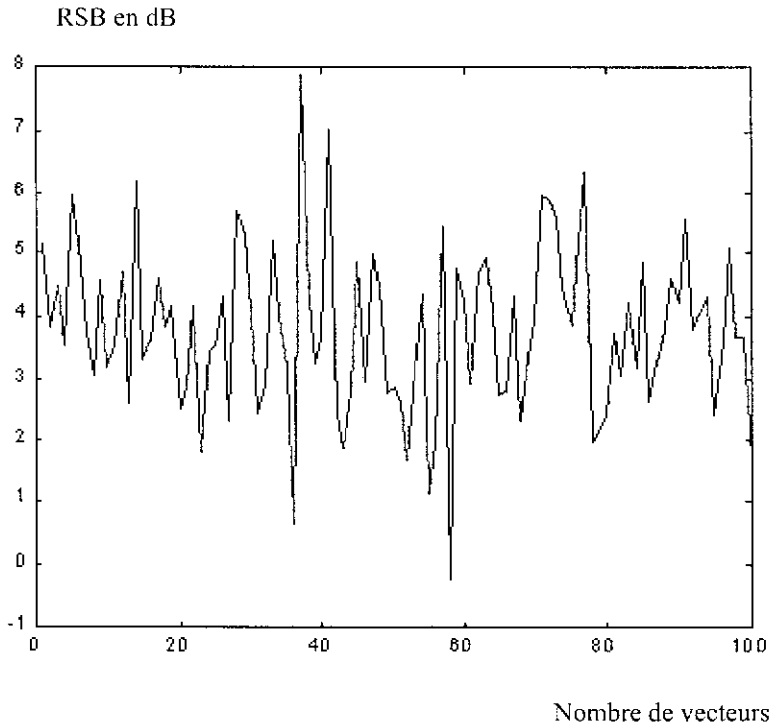


Fig 9.4. Rapport signal sur bruit pour un dictionnaire de taille 128 points.

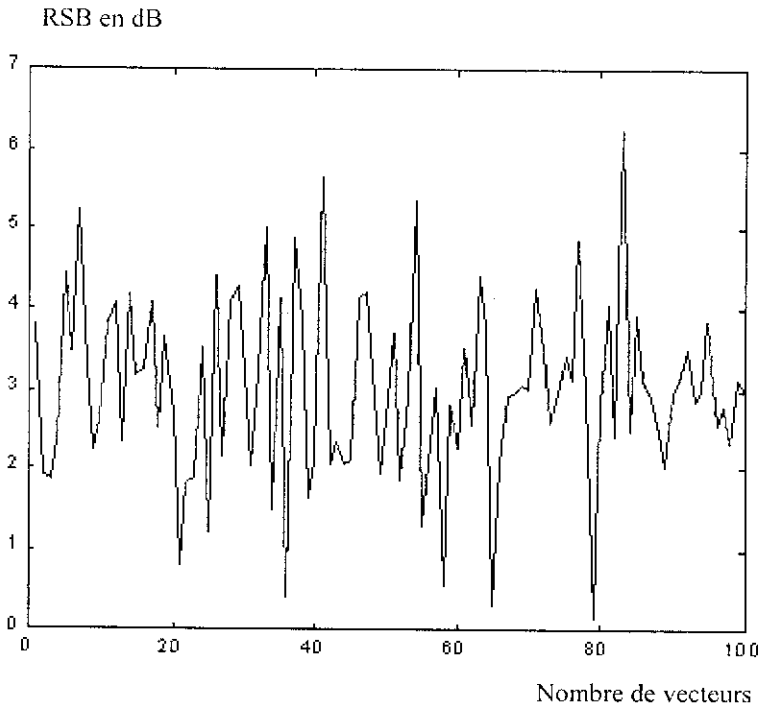


Fig 9.5. Rapport signal sur bruit pour un dictionnaire de taille 64 points.

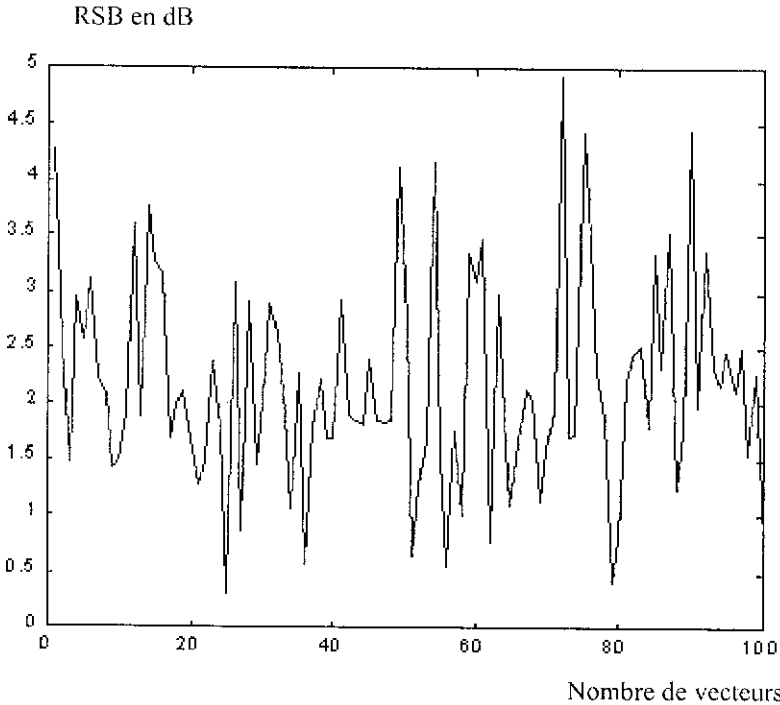


Fig 9.6. Rapport signal sur bruit pour un dictionnaire de taille 32 points.

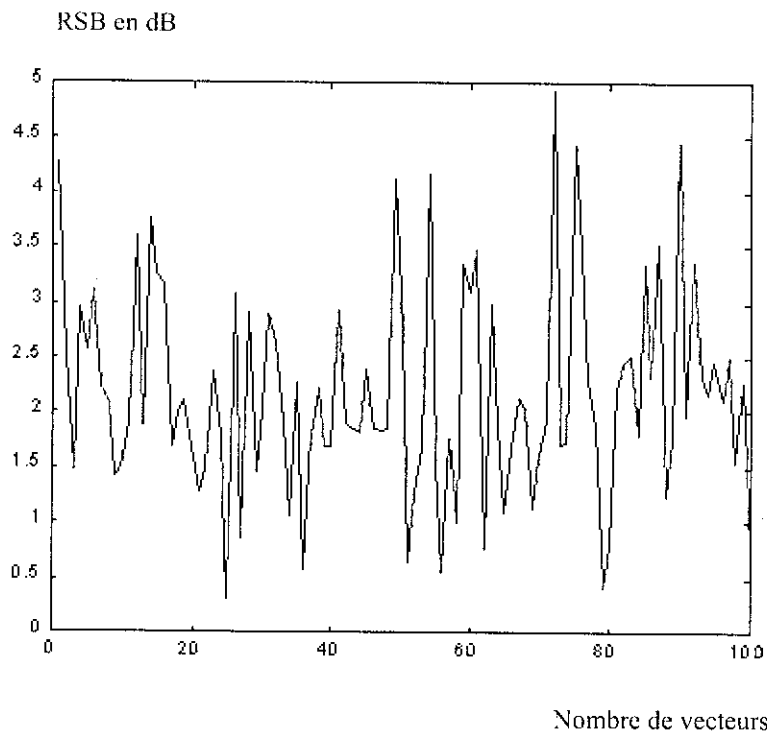


Fig 9.7. Rapport signal sur bruit pour un dictionnaire de taille 16 points.

Les performances du quantificateur utilisé sont présentées dans la table (9.1). En effet ces résultats montrent bien que le rapport signal sur bruit croit avec la taille du dictionnaire.

Taille	512	256	128	64	32	16
RSB(dB)	4.31	4.31	3.78	2.99	2.17	1.66

Tab 9.1. Tableau représentatif des valeurs RSB pour différentes tailles de dictionnaire

10. Expérimentation et résultats

Dans ce qui suit nous présentons les résultats émanant des expériences faites sur la compression du signal parole en utilisant la transformation SVD comme illustré dans le schéma (8.1).

Les paramètres d'entrée de notre système de codage sont les coefficients de prédiction linéaire $a(i)$. Ces derniers sont extraits au nombre de dix pour chaque trame du signal vocal, l'ordre de prédiction choisi étant égal à dix.

Les $a(i)$ représentent les éléments de la matrice A , sur laquelle nous avons appliqué la SVD. La matrice V obtenue est notre matrice de transformation.

La quantification vectorielle est appliquée ensuite sur les coefficients transformés. Pour cela nous avons construit au préalable un dictionnaire de vecteurs prototypes ayant une taille de 128 vecteurs, en utilisant l'algorithme LBG.

A la réception, pour restituer les coefficients $\hat{a}(i)$, nous avons utilisé la formule :

$$A = USV^T$$

En excitant par le résidu, le filtre de synthèse, dont les paramètres sont les coefficients $\hat{a}(i)$ susmentionnés, nous avons reconstruit le signal codé.

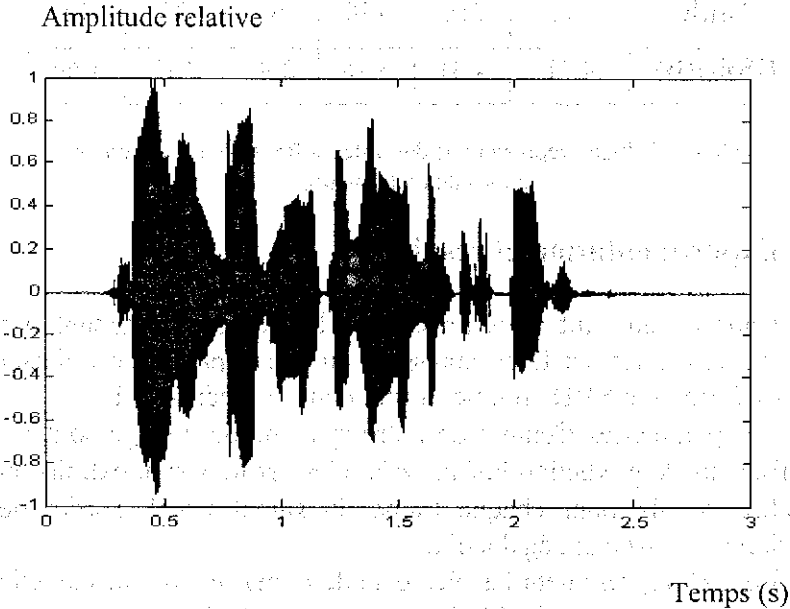


Fig 10.1. Le signal vocal correspondant à la phrase avant l'opération de compression.

Nous avons utilisé plusieurs phrases lors du test. Toutefois, en raison de la similitude des résultats, nous nous sommes contentés de montrer ceux concernant une phrase d'une durée d'environ trois secondes - représentée par la figure (10.1) - ainsi que ceux d'un segment de parole en l'occurrence la voyelle « a » - figure (10.2) -.

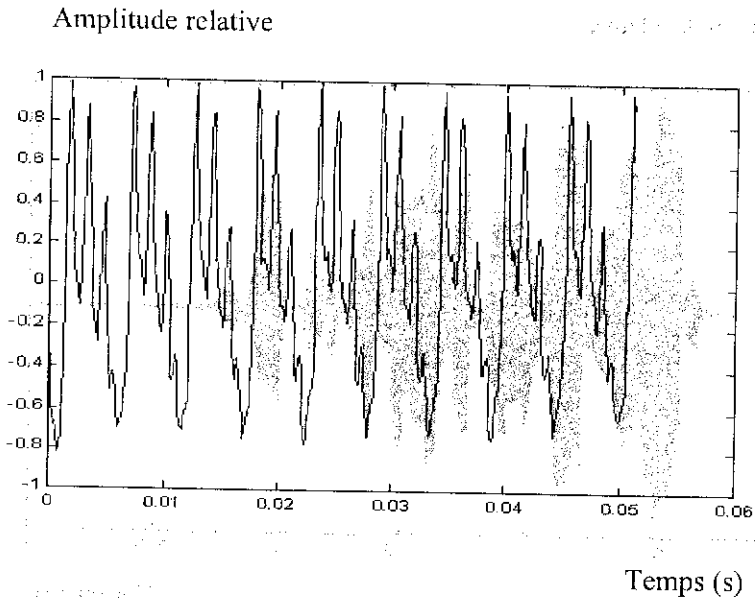


Fig 10.2: Le signal vocal correspondant à la voyelle 'a' avant l'opération de compression.

La figure (10.3) représente l'onde temporelle restituée, correspondante à la phrase. La figure (10.4) montre celle du son voisé "a". En les comparant avec les signaux originaux présentés respectivement dans les figures (10.1) et (10.2) on peut noter la grande ressemblance existante entre les divers signaux (originaux et synthétiques).

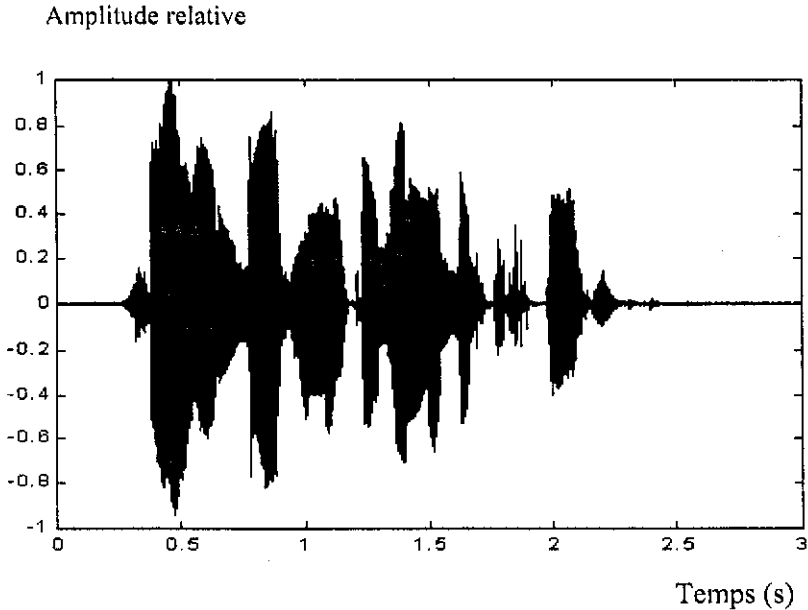


Fig 10.3. Le signal correspondant à la phrase, restitué après l'opération de compression.

Toutefois cette similitude n'est pas toujours nécessaire pour avoir une bonne qualité synthétique du signal codé [10,11]. Pour mesurer la qualité de ce dernier, des tests subjectifs sont sollicités.

La méthode MOS (Mean Opinion Score) est l'une des méthodes de test les plus utilisées. Elle consiste à réaliser l'écoute des signaux par un certain nombre d'individus (allant de 12 à 24).

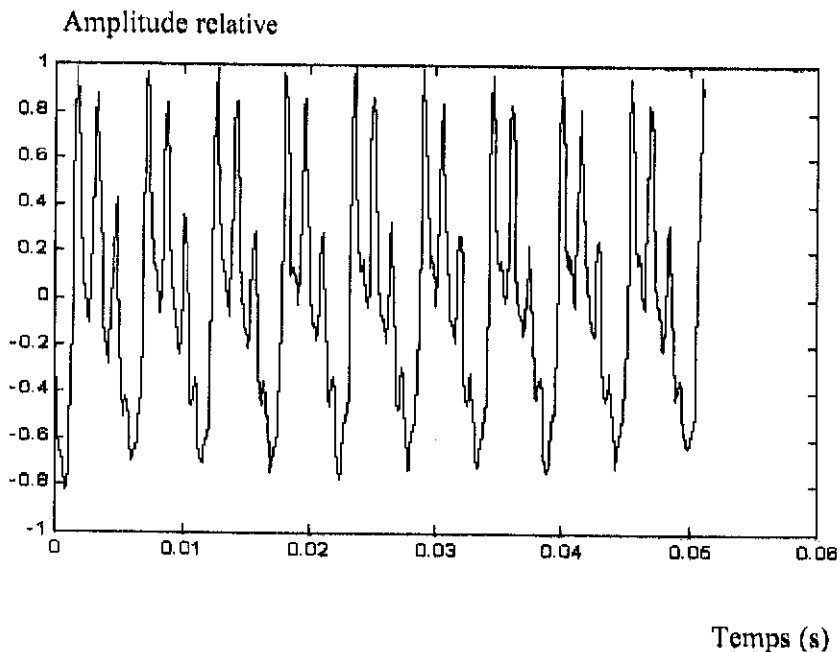


Fig 10.4. L'onde temporelle de la voyelle "a" restituée après la compression.

Des notes sont ensuite données par ces derniers, suivant les dégradations perçues. L'échelle de notation sur laquelle se basent ces mesures subjectives varie entre 5 et 1, où la note 5 indique que les dégradations sont imperceptibles tandis que la note 1 révèle la présence d'un bruit très gênant. Concernant les phrases utilisées pour le test, la note qui a été attribuée est 5.

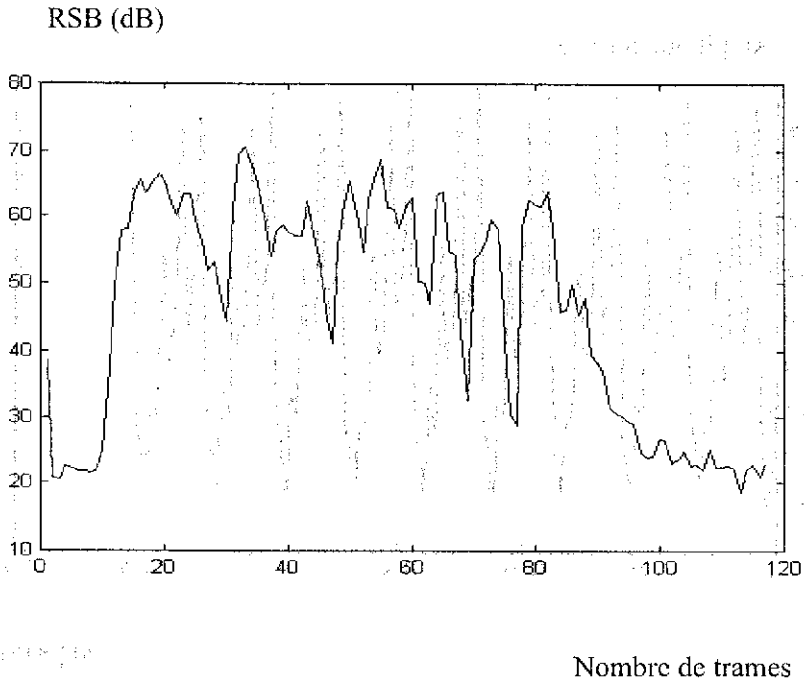


Fig 10.5. Le rapport signal sur bruit
(signal correspondant à la phrase)

La figure (10.5) confirme la ressemblance apparente qui existe entre les signaux originaux et ceux restitués, en montrant les valeurs du rapport signal sur bruit obtenu par la courbe tracée. La valeur moyenne du RSB est égale à 46.55 dB.

L'espérance des rapports signal sur bruit pour tous les signaux de parole que nous avons utilisés avoisine la valeur 46 dB.

11. Conclusion :

En écoutant les phrases après leurs restitutions, la qualité de la synthèse a été jugée satisfaisante par plusieurs individus. Il n'a été décelé aucun bruit gênant.

La compression par transformation a été envisagée pour réduire le débit, dans le sens où elle concentre l'énergie du signal dans un nombre faible de coefficients. L'utilisation de la SVD ou d'une autre transformation telle que la DCT est en général préférée pour cette raison.

L'intérêt du codage par transformation réside dans la quantification, avec plus de résolution, pour certains coefficients transformés que pour d'autres, c'est l'allocation dynamique de bits.

En utilisant le codage sans transformation, on octroie les bits d'une façon équivalente aux différentes composantes du vecteur à quantifier. Par contre dans le cas du codage par transformation, on alloue plus de bits aux coefficients qui contiennent plus d'énergie et on néglige ceux de faible valeur (tel est le cas de la SVD). De ce fait apparaît l'importance du codage par transformation où l'on consomme moins de bits.

Ce que nous n'avons pas réalisé dans ce travail c'est l'allocation dynamique de bits, qui augmente considérablement l'efficacité du codage par transformation. C'est ce que nous projetons de réaliser dans l'avenir.

Bibliographie

المراجع

- [1] M. Abbas, "Application de la transformée orthogonale SVD en compression de la parole", thèse de magister, CRSTDLA-ENS, 2002.
- [2] R. Boite, M. Kunt, "Traitement automatique de la parole", Presses Polytechniques Romandes, Lausanne, Suisse, 1987.
- [3] P. Noll, N.S. Jaynt, "Digital Coding Of Waveforms", Pr-Hall, 1984.
- [4] S.K Aoubid, M Boulemden, "Compression du signal de parole par les transformations de Fourier, Cosinus, Hadamard, et en Ondelettes", Proceedings of the IEE'97, International Conference, Vol. Electronic Engineering, 7-9 Dec. 1997, Batna University, Algeria.
- [5] D. Berkani, "Application de la transformée orthogonale SVD en compression de la parole", Séminaire National sur l'Automatique et les Signaux SNAS ' 99, (1999), Annaba University, pp. 132-135.
- [6] L. Knockaert, B. De Backer, and D. De Zutter, "SVD compression, Unitary Transforms, and Computational Complexity", IEEE Transactions on Signal Processing, Vol 47, NO 10, October 1999.
- [7] E. Biglieri, K. Yao, "Some Properties of singular value decomposition and their applications to digital signal processing", Signal Processing Vol. 18 Nov 1989.
- [8] A. Gersho, R. Gray, "Vector quantization and signal compression", Kluwer Academic Publishers, Boston / Dordrecht / London 1992, sixth printing 1997.
- [9] Calliope, "la parole et son traitement automatique", Editeur J.P.Tubach, Masson, Paris, 1989.
- [10] M. Abbas, D. Berkani, "Le codage de la parole par les deux transformées SVD et DCT," Rencontres Jeunes Chercheurs en parole RJC ' 2003, du 23 au 25 septembre 2003, Grenoble, France.
- [11] M. Abbas, D. Berkani, "L'utilisation de la transformation SVD dans la compression des signaux de parole," Séminaire National sur l'Automatique et les Signaux SNAS ' 02, 27 et 28 octobre 2002 Université de Annaba, Annaba, Algérie.