

EGALISATION ET DECODAGE A MAXIMUM DE VRAISEMBLANCE SYMBOLE PAR SYMBOLE  
APPLICATION DE L'ALGORITHME DE BAUM-WELCH

Robert VALLET

TELECOM PARIS 46 rue Barrault 75634 PARIS, FRANCE

RESUME

Cet article décrit un démodulateur composé d'un égaliseur à maximum de probabilité a Posteriori symbole par symbole (MAPS), associé à l'algorithme d'identification des chaînes de Markov cachées de Baum-Welch. L'identification des paramètres du canal de transmission est réalisée sans séquence d'apprentissage par un algorithme itératif. Un décodeur à Maximum de Probabilité a Posteriori symbole par symbole réalise le décodage à décisions souples d'un code convolutif. Le gain en rapport signal sur bruit, est de l'ordre de 3 dB, par rapport à un égaliseur de Viterbi suivi d'un décodeur à décisions dures.

ABSTRACT

This paper deals with a blind equalizer based on the Maximum a posteriori symbol by symbol (MAPS) criteria, associated with the Baum-Welch identification algorithm of Hidden Markov Models. The channel parameters are estimated without the presence of a training period. A MAPS decoder is used to obtain a soft decision decoding procedure of a convolutional code. The gain in signal to noise ratio is on the order of 3 dB in comparison to a Viterbi equalizer and a Hard decision decoder.

Introduction:

L'algorithme de Viterbi V.A. [VIT71], [FOR73], est devenu l'algorithme indispensable à la réalisation des récepteurs en communications numériques. Il réalise les fonctions de démodulation, d'égalisation et de décodage, suivant le critère de Maximum de Vraisemblance (ML) ou de Maximum de Probabilité a Posteriori (M.A.P.) de la suite des symboles d'information. Un entrelaceur permet de décorrélérer les erreurs présent à la sortie de l'égaliseur. Le décodeur est à décisions dures. L'ensemble est représenté sur la figure 1-a. L'égaliseur à Maximum de probabilité a posteriori symbole par symbole [FOR73], [VAL89] est réalisé par un algorithme Forward-Backward de lissage non linéaire. Un décodeur à décisions souples, basé sur le même algorithme de lissage non linéaire, permet d'évaluer la probabilité a posteriori des symboles d'information. L'algorithme de Baum-Welch est associé au lisseur non linéaire pour identifier les paramètres du canal de transmission sans séquence d'apprentissage. Le schéma de principe est représenté sur la figure 1-b.

Cet article a pour objectif d'apporter des solutions optimales à la mise en cascade d'un égaliseur et d'un décodeur à maximum de vraisemblance symbole par symbole et à associer, à l'égaliseur, un algorithme d'identification optimal du canal de transmission conformément à la figure 1-b.

Egalisation a maximum de vraisemblance :

Représentation de l'observation :

L'observation est créée par la transmission d'une modulation numérique linéaire dans un canal dispersif à bruit Additif Gaussien Blanc AGB.

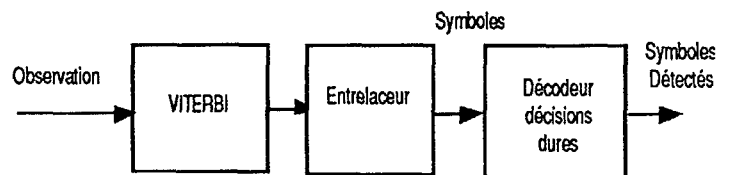


Figure 1a Egalisation par l'algorithme de Viterbi, Entrelacement des symboles et décodage à décisions dures.

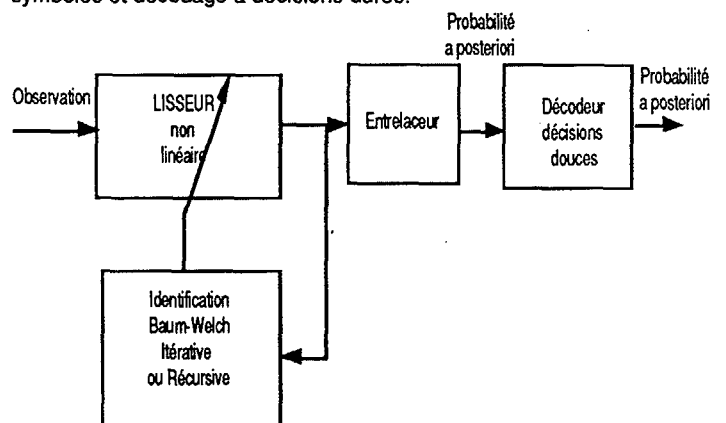


Figure 1-b Egalisation par l'algorithme de lissage non linéaire, Identification du canal de transmission, Entrelacement des probabilités a posteriori des symboles et décodage à décisions douces.

Elle est modélisée par une chaîne de Markov à temps discret et à états discrets. Le modèle d'états est décrit par l'équation d'états.

$$x_{n+1} = A x_n + B d_n \quad (1)$$

$$\text{où } x_{n+1} = (d_n, d_{n-1}, \dots, d_{n-L+2})^T \text{ et } x_n = (d_{n-1}, d_{n-2}, \dots, d_{n-L+1})^T \quad (2)$$

$d_n$  représente le symbole d'information transmis à l'instant  $n$ ,  $A$  est une matrice de décalage et  $B$  un vecteur  $B^T = (1, 0, 0, 0, \dots, 0)$ . La transition  $\xi_n$



entre deux états est définie par

$$\xi_n = (d_n, d_{n-1}, \dots, d_{n-L+1})^T = (x_{n+1}^T, d_{n-L+1})^T = (d_n, x_n^T)^T \quad (3)$$

L'équation d'observation est définie par:

$$y_n = h(\xi_n) + b_n \quad (4)$$

La loi de l'observation  $y_n$  conditionnellement à une transition  $\xi_n$  s'exprime par une fonction Gaussienne de moyenne  $h(\xi_n)$  et de matrice de covariance  $N_0 I_\mu$  ou  $\xi_n$  représente une transition possible du treillis associé au modèle et  $\mu$  le nombre d'échantillons du signal observés pendant la durée d'un symbole d'information. Si le canal de transmission est linéaire, la contribution  $h(\xi_n)$  de l'état à la sortie du système est définie à partir de la réponse impulsionnelle du canal et de la branche  $\xi_n$  par une relation de convolution:

$$y_n = \sum_{k=0}^{L-1} h_k d_{n-k} + b_n = H^T \xi_n \quad (5)$$

La métrique de branche est définie par la probabilité du couple  $(x_{n+1}, y_n)$  conditionnellement à l'état  $x_n$  à l'instant  $n$ :

$$m_\lambda(\xi_n) = P_\lambda(x_{n+1}, y_n | x_n) \quad (6)$$

L'ensemble des paramètres du modèle est défini par  $\lambda = (h(\xi_n), \sigma^2)$ , il caractérise la structure du treillis et la métrique de branche associée.

#### Egalisation à Maximum de vraisemblance symbole par symbole:

L'évaluation de la probabilité a posteriori des symboles transmis est réalisée par un lisseur non linéaire normalisé, le vecteur  $\lambda$  est supposé connu de l'utilisateur. Cet algorithme est composé d'un filtre prédicteur non linéaire direct et d'un filtre retour non linéaire, qui permettent d'évaluer la probabilité a posteriori de chaque état du modèle. La probabilité a posteriori des symboles d'information est obtenue par sommation partielle de celle des états ou des transitions. Le lisseur non linéaire est constitué du filtre prédicteur direct normalisé :

$$P_\lambda(x_n | y_1^{n-1}) = \sum_{d_{n-L+1}=1}^M \exp \frac{|y_{n-1} - h(\xi_{n-1})|^2}{2N_0} P_\lambda(x_{n-1} | y_1^{n-2}) / N_{\alpha n-1} \quad (7)$$

et du filtre non linéaire retour "normalisé" :

$$\beta'_\lambda(x_n | y_n) = \sum_{x_{n+1}=1}^{M'} \exp \frac{|y_n - h(\xi_n)|^2}{2N_0} \beta'_\lambda(x_{n+1} | y_{n+1}^N) / N_{\alpha n} \quad (8)$$

La probabilité a posteriori de l'état  $x_n$  est définie par le produit du prédicteur direct et du filtre retour:

$$P_\lambda(x_n | y_1^N) = P_\lambda(x_n | y_1^{n-1}) \beta'_\lambda(x_n | y_n^N) \quad (9)$$

La probabilité a posteriori d'une transition est définie par :

$$P_\lambda(\xi_n | y_1^N) = P_\lambda(x_n | y_1^{n-1}) \exp \frac{|y_n - h(\xi_n)|^2}{2N_0} \beta'_\lambda(x_{n+1} | y_{n+1}^N) / N_{\alpha n} \quad (10)$$

La probabilité à posteriori du symbole  $d_n$  est évaluée par sommation partielle de celle des états :

$$P_\lambda(d_n | y_1^N) = \sum_{(d_{n-1}, d_{n-2}, \dots, d_{n-L+2})} P_\lambda(x_{n+1} | y_1^N) \quad (11)$$

Cette probabilité a posteriori sera transmise au décodeur à entrée pondérée ou directement au système de décision s'il n'y a pas de codage correcteur d'erreurs. La décision à maximum de vraisemblance du symbole transmis à l'instant  $n$  est alors réalisée par:

$$d_{jn} \text{ tel que } P_\lambda(d_{jn} | y_1^N) > P_\lambda(d_{in} | y_1^N) \text{ quelque soit } j \neq i \quad (12)$$

Un algorithme récursif par bloc, de complexité finie est utilisable lorsque le nombre d'observations  $N$  est grand devant le retard de décision,  $D = 10L$ , associé à un égaliseur de Viterbi. Cet algorithme permet également d'obtenir un retard de décision borné  $2D$ . Il est utilisable dans un système de transmission en temps réel.

#### Identification du canal de transmission : Algorithme de Baum-Welch.

L'identification des paramètres du canal de transmission est réalisée par l'algorithme de BAUM WELCH (BW) [BAU73], [RAB86], [VAL89]. Le canal est dispersif à bruit additif Gaussien blanc et indépendant de l'état  $x_n$  du modulateur. Les relations d'identification itératives sont obtenues à partir de la probabilité a posteriori des transitions du treillis (10)  $P_\lambda(\xi_n = i, j | y_1^N)$  L'identification des valeurs moyennes  $h(\xi_n)$  de la sortie est réalisée par la relation :

$$h'(\xi_n) = \frac{1}{N} \sum_{n=1}^N P_\lambda(\xi_n | y_1^N) y_n \quad \text{Pour } \xi_n = i, j, i, j = 1, \dots, M'. \quad (13)$$

La variance du bruit est estimée par la relation :

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^N \sum_{\xi_n} P_\lambda(\xi_n | y_1^N) |y_n - h(\xi_n)|^2 \quad (14)$$

La relation (13) permet d'identifier les paramètres d'un canal non-linéaire AGB. En général, le modulateur et le canal de transmission sont tous les deux linéaires. Ils sont caractérisés par une réponse impulsionnelle de durée finie qu'il faut identifier. Des relations particulières d'identification sont alors disponibles.

#### Identification de la réponse impulsionnelle du canal:

La réponse impulsionnelle du canal de transmission est estimée par la solution de la relation matricielle [VAL91]:

$$R_{y,d} = R_{d,d} \cdot H' \quad (15)$$

La matrice d'intercorrélation des données  $R_{d,d}$  a pour élément

$$r_{d,d}(i,j) = \sum_{n=0}^N \sum_{d_{n-j}, d_{n-i}} P_\lambda(d_{n-j}, d_{n-i} | y_0^N) d_{n-i} d_{n-j} \quad (16)$$

Le vecteur d'intercorrélation données-observations  $R_{y,d}$  a pour élément :

$$r_{y,d}(i) = \sum_{n=0}^N \sum_{d_{n-j}} P_\lambda(d_{n-j} | y_0^N) y_n d_{n-j} \quad (17)$$

L'inter-corrélation symboles données et l'auto-corrélation des symboles sont évaluées de façon "douce" à partir des probabilités a posteriori empiriques des différentes variables.

#### Entrelacement:

Un entrelaceur convolutif [RAM71] de paramètres  $n_1=11$  et  $n_2=12$  est inséré entre le codeur et le modulateur. Cet entrelaceur apporte un retard global de 121 durées symboles.

#### Décodage pondéré des codes convolutifs:

L'égaliseur par lissage non linéaire fournit à l'entrée du décodeur les probabilités a posteriori des symboles non codés  $P(d_n | Y)$  pour  $n=1, \dots, N$ , où  $d_n$  est une variable à  $M$  valeurs non directement accessible à l'observateur, qui représente les symboles codés. L'entrelaceur rend ces variables aléatoires  $d_n$  statistiquement indépendantes. Pour ne pas alourdir encore les notations, les mêmes notations indicielles, à la sortie et à l'entrée de l'entrelaceur, sont utilisées. Le décodage pondéré est

réalisé à partir de l'observation de ces probabilités conditionnelles.

**Représentation d'un codeur convolutif :**

Le codeur convolutif de taux  $k/\mu$  et de longueur de contrainte  $L$ , est représenté par une équation d'état:

$$X_{n+1} = S X_n + A_n \tag{18}$$

et une équation d'observation:

$$S_n = G \xi_n \tag{19}$$

L'état est défini par  $X_n = (A'_{n-1}, A'_{n-2}, \dots, A'_{n-L+1})'$  à l'instant  $n$  et  $X_{n+1} = (A'_n, A'_{n-1}, \dots, A'_{n-L+2})'$  à l'instant  $n+1$  et  $\xi_n = (A'_n, A'_{n-1}, \dots, A'_{n-L+1})' = (X_n, X_{n+1})'$  est la transition entre deux états consécutifs.

L'état est un vecteur de dimension  $(L-1)k$ , La transition  $\xi_n$  est un vecteur de dimension  $Lk$ .  $S$  est une matrice de décalage d'ordre  $k$  de dimension  $(L-1)k$  et  $A_n = (a_{nk}, a_{nk+1}, \dots, a_{nk+k-1}, 0, 0, \dots, 0)$  est un vecteur de dimension  $(L-1)k$  dont les  $(L-2)k$  dernières composantes sont nulles. Les  $k$  premières composantes représentent les  $k$  symboles d'entrée du codeur.  $S_n = (d_{\mu n}, d_{\mu n+1}, \dots, d_{\mu n+\mu-1})^T$  est la suite des valeurs des symboles codés associés à une transition  $\xi_n$ . Un codeur convolutif de taux  $1/2$  et de longueur de contrainte  $L=3$  est représenté sur la figure (2).

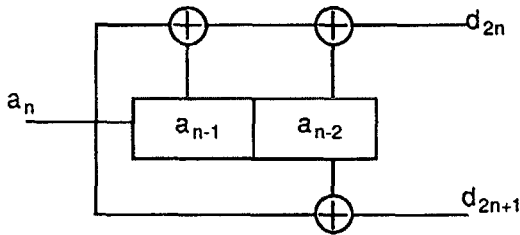


Figure 2: Codeur convolutif de taux  $1/2$  et de longueur de contrainte  $L=3$ .

**Décodeur à lissage non linéaire :**

L'observation, à l'entrée du décodeur, est définie par la probabilité a posteriori de la suite  $(d_1, d_2, \dots, d_N)$  qui s'exprime par le produit des probabilités a posteriori des symboles codés obtenues par l'égaliseur à lissage non linéaire :

$$P(d_1, \dots, d_N | Y) = \prod_{n=1}^N P(d_n | Y) \tag{20}$$

$Y = \{y_1^N\}$  représente le signal observé à l'entrée de l'égaliseur. La

relation précédente est obtenue par l'action de l'entrelaceur et par le fait que le décodeur ne peut pas prendre en compte la corrélation qui existe entre des symboles trop éloignés, comparativement à la longueur de contrainte du code. La probabilité a posteriori d'une suite  $S_n = (d_{\mu n}, d_{\mu n+1}, \dots, d_{\mu n+\mu-1})^T$  est définie à partir de (19) par la relation :

$$P(G \xi_n | Y) = \prod_{m=0}^{\mu-1} P(d_{n\mu+m} | Y) \tag{21}$$

On n'a pas accès à l'observation, comme pour un modèle de Markov caché, mais seulement à sa probabilité conditionnellement à une transition  $\xi_n$ .

La probabilité d'une trajectoire  $X = \{x_0^N\}$ , conditionnellement à l'observation  $Y = \{y_0^N\}$  est définie par:

$$P(X|Y) = P(x_0) \prod_{n=1}^N P(x_{n+1}|x_n) P(G \xi_n | Y) \tag{22}$$

La métrique associée à la transition  $\xi_n$  est définie par :

$$m(\xi_n) = P(x_{n+1}|x_n) P(G \xi_n | Y) = \prod_{m=0}^{\mu-1} P(d_{n\mu+m} | Y) \text{ si } \xi_n \text{ existe} \tag{23}$$

$$m(\xi_n) = 0 \text{ sinon.}$$

A partir de la définition d'un modèle de Markov à états discrets (18), (19) et d'une métrique de branche (23), l'algorithme de Viterbi détermine la trajectoire de probabilité a posteriori maximale. L'algorithme de lissage non-linéaire évalue la probabilité à posteriori de chaque états et de chaque symbole d'information  $a_n$ .

**Lissage non linéaire:**

L'algorithme de lissage non linéaire est défini à partir de la métrique de branche définie par la relation (23). Le filtre prédicteur non linéaire normalisé est défini par :

$$P(x_{n+1}|Y_0^N) = \sum_{a_{n-L+1}} P(x_{n+1}|x_n) P(G(\xi_n) | Y) P(x_n|Y_0^{n-1}) / N_{\alpha n} \tag{24}$$

Le filtre retour "normalisé" est défini par:

$$\beta'(x_{n+1}|Y_N^{n+2}) = \sum_{a_{n+1}} P(x_{n+2}|x_{n+1}) P(G(\xi_{n+1}) | Y) \beta'(x_{n+2}|Y_N^{n+3}) / N_{\alpha n+1} \tag{25}$$

La probabilité a posteriori d'un état  $x_{n+1}$  est définie par le produit:

$$P(x_{n+1}|Y) = P(x_{n+1}|Y) \beta'(x_{n+1}|Y) \tag{26}$$

La probabilité a posteriori des symboles d'information est évaluée à partir du lisseur non linéaire par:

$$P(a_n|Y_0^N) = \sum_{a_{n-1}, \dots, a_{n-L+2}} P(x_{n+1}|Y_0^N) \tag{27}$$

La décision de la suite de symboles d'information, suivant le critère de Maximum de Probabilité a posteriori, est définie par:

$$\hat{a}_n \text{ tel que } P(\hat{a}_n|Y_0^N) > P(a_n|Y_0^N) \quad \forall a_n \neq \hat{a}_n \tag{28}$$

Les relations précédentes permettent de propager à un décodeur extérieur la probabilité a posteriori des symboles (27). Les décodeurs symbole par symbole peuvent être concaténés sans difficultés particulières, en plaçant un entrelaceur de longueur appropriée.

**Résultats de simulations:**

Les deux algorithmes d'égalisation de Viterbi VA et de lissage non linéaire LNL ont été simulés sur un canal de transmission, à temps discret, de réponse impulsionnelle 0,5, 0,7 et 0,5.

**Identification du canal de Transmission:**

L'identification des paramètres a été réalisée pour une observation de  $N=64$  symboles d'information. Les valeurs estimées des coefficients de la réponse impulsionnelle par l'algorithme de Baum-Welch (15) à (17), en fonction de l'indice de l'itération sont représentées sur la figures 3 pour un  $E_b/N_0$  de 4dB et des valeurs initiales de la réponse impulsionnelle  $h(0)=0, h(1)=1$  et  $h(2)=0$ . L'identification est presque réalisée avec une seule itération.

**Variance de l'erreur d'identification:**

Une façon de caractériser l'erreur d'identification est d'étudier l'erreur quadratique d'estimation, qui est définie comme la distance quadratique moyenne entre les valeurs exactes  $h_0(x_n)$  de la sortie du canal et celles estimées  $h(x_n)$ . Elle est définie par:

$$\epsilon^2 = \frac{1}{M} \sum_{\xi_n=1}^M (h_0(\xi_n) - h(\xi_n))^2 \tag{29}$$

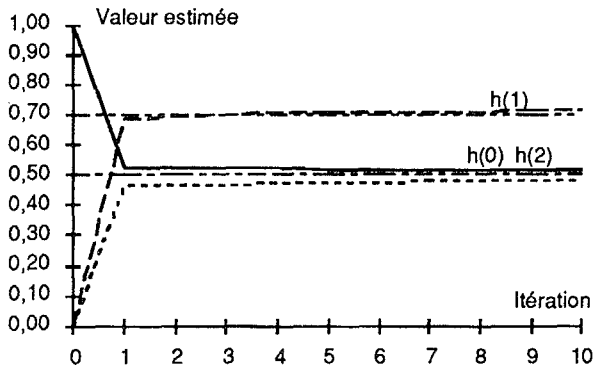


Figure 3 Réponse impulsionnelle estimée avec l'algorithme de Baum-Welch  $E_b/N_0 = 4 \text{ dB}$   $N = 64$  Observations, Initial 1,0,0

Les résultats de simulations obtenus sont représentés sur la figure 4. L'identification du canal de transmission avec une séquence d'apprentissage sert de référence. L'identification des coefficients de la réponse impulsionnelle du canal par l'algorithme de Baum-Welch (15) à (17) produit une erreur quadratique équivalente à la précédente. L'identification des valeurs des sorties du canal par l'algorithme (13) est moins performante, mais elle pourrait permettre d'identifier un canal non linéaire. L'algorithme de Viterbi permet de déterminer une suite de symboles qui est utilisée pour l'identification du canal à la place de la séquence d'apprentissage, l'erreur d'identification est un peu plus grande.

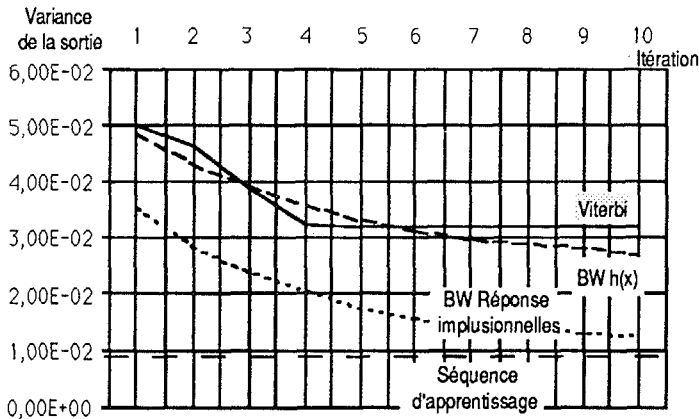


Figure 4: Comparaison des erreurs quadratique d'identification pour les trois algorithmes de Viterbi, BW sur les sorties et BW sur la réponse impulsionnelle, pour  $E_b/N_0 = 4 \text{ dB}$ .

Les performances obtenues à partir du schéma 1-b constitué d'un égaliseur, d'un entrelaceur et d'un décodeur de codes convolutif sont représentées pour le canal de transmission 0,5, 0,7, 0,5 et le codeur de la figure 2. La figure 5 regroupe l'ensemble des performances en Taux d'erreur par élément binaire TEB.

-> Les TEB obtenus à la sortie de l'égaliseur avec un algorithme de Viterbi ou de lissage non linéaire sont équivalents.

-> La réalisation d'un décodeur pondéré à la sortie du lisseur non linéaire, figure 1-b permet d'obtenir un gain significatif de l'ordre de 3dB par rapport à l'utilisation d'un décodeur à décisions dures après un égaliseur de Viterbi figure 1-a.

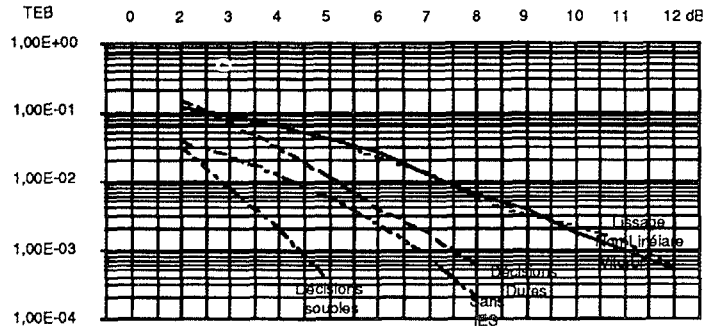


Figure 5: Performances obtenues pour un canal 0.5, 0.78, 0.5, un entrelaceur 11,12 et un codeur L=3 de taux 1/2. Comparaison des égaliseurs et du décodeur à décisions souples par rapport au décodeur à décisions dures.

**Conclusion :**

L'égaliseur à Maximum de probabilité a Posteriori symbole par symbole ou de lissage non linéaire, permet d'utiliser l'algorithme de Baum-Welch, dont il est la première étape, pour identifier les paramètres du canal de transmission avec une erreur quadratique d'identification équivalente à celle obtenue avec une séquence d'apprentissage. Il permet aussi de réaliser un décodeur pondéré par un algorithme de lissage non linéaire. Un gain de l'ordre de 3 dB en rapport signal sur bruit par comparaison avec l'utilisation d'un décodeur à décisions dures est obtenu.

**Références :**

[VIT71] A.J.VITERBI. "Convolution Codes and Their Performances in Communications Systems".IEEE Transactions on Communications, VOL. COM-19, No.5, October 1971.

[FOR73] FORNEY,"The Viterbi Algorithm".IEEE Transaction on Information Theory, Vol.IT-19, No.2, March 1973.

[BAU73] L.E.BAUM. "An Inequality and Associated Maximisation Technique in Statistical Estimation for Probabilistic Functions of Markov Processes".INEQUALITIES, VOL. III Academic press New York 1972

[RAB86] L.R.RABINER and B.H.JUANG. " An Introduction to Hidden Markov Models". IEEE ASSP Magazine, JANUARY 1986.

[VAL89] R.VALLET et H.KOREZLIOGLU. "Identification des chaînes de Markov cachées: Applications aux canaux non-linéaires".12-eme Colloque GRETSI, JUAN-LES-PINS, MAI 1989.

[RAM70] J.L.RAMSEY. "Realization of Optimum Interleaver". IEEE Transactions on Information Theory,VOL.IT-16,No.3, Mai 1970

[VAL91] R.VALLET and G.KAWAS KALEH "Joint Channel Identification and Symbols Detection " IEEE International Symposium on Information Theory Budapest 1991.

[KAW89] G.KAWAS KALEH et R.VALLET "Détection et estimation pour la transmission sur un canal inconnu".12-eme Colloque GRETSI, JUAN-LES-PINS, MAI 1989