

Un prétraitement générique pour les STANN

Illustration en lecture labiale

David Mercier, Renaud Séguier

Supélec : Équipe Electronique Traitement du Signal et Neuromimétisme

Avenue de la Boulaie - BP28 35511 Cesson Sévigné, France.

e-mail : David.Mercier@supelec.fr, Renaud.Segulier@supelec.fr

Résumé

Notre équipe travaille depuis une dizaine d'années sur un modèle de réseaux de neurones utilisant non pas des entrées continues mais des entrées impulsionnelles, s'inspirant en cela de la nature des entrées des neurones biologiques. Ce modèle, appelé STANN (Spatio-Temporal Artificial Neural Network), permet ainsi de traiter, comme son nom l'indique, des données spatio-temporelles, c'est à dire des données où l'information spatiale évolue au cours du temps. La mise en pratique de cette famille de réseaux de neurones sur les problèmes de l'écriture manuscrite [11] et de la lecture labiale [2] en a montré le potentiel. Mais elle a également mis en avant un problème courant dans les réseaux de neurones à impulsion : la difficulté de générer des signaux impulsionnels à partir des signaux spatio-temporels (séquence d'images, signaux multicapteurs, etc.) constituant les entrées. Nous présentons dans cet article une méthode simple et générique pour créer ces signaux impulsionnels depuis une base de signaux évoluant dans le temps de façon continue. Cette méthode est basée sur la quantification vectorielle. Nous l'avons appliquée à la lecture labiale et avons ainsi pu comparer les résultats avec ce prétraitement aux résultats obtenus auparavant ([1] et [2]).

1 Les STANN

1.1 Modélisation du neurone : le STAN

Le STAN (Spatio-Temporal Artificial Neuron) est un modèle de neurone créé par Vaucher [15], le codage sous-jacent ayant été intégré dans les architec-

tures neuronales classiques ([10]). Son principe est de coder des événements discrets ayant deux degrés de liberté (amplitude et date) sous la forme de nombre complexes ayant aussi deux degrés de liberté (amplitude et phase).

Un neurone STAN est caractérisé par au plus cinq éléments (voir Figure 1). Tout d'abord, comme pour le modèle neuronal classique, un STAN se singularise des autres par son *vecteur poids* (W), sa *fonction potentielle* (V ou D) et sa *fonction de transfert* (ou *fonction d'activation* F). Eventuellement, mais nous ne l'avons pas représenté sur la Figure 1, un quatrième paramètre peut être défini : le *biais* (b). Le dernier paramètre, nommé TW , représente la taille de la fenêtre temporelle à l'intérieur de laquelle on désire identifier des séquences d'impulsions (on peut l'assimiler au retard maximal dans un neurone classique dynamique).

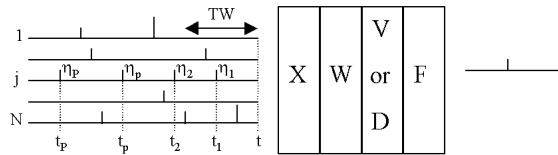


FIG. 1 – Le STAN (Spatio-Temporal Artificial Neuron)

Les calculs se mènent de la façon suivante : une impulsion d'amplitude η_1 émise à l'instant t_1 sur la $j^{\text{ème}}$ composante du vecteur d'entrée X est codée à l'instant courant t par le nombre complexe :

$$x_j(t) = \eta_1 e^{-\mu_S \tau_1} e^{i \arctan \mu_T \tau_1}$$

$$i = \sqrt{-1}, \tau_1 = t - t_1 \text{ et } \mu_S = \mu_T = 1/TW$$

Si une seconde impulsion d'amplitude η_2 est émise à l'instant t_2 sur la même entrée, elle est ajoutée à l'état courant et c'est ce résultat que l'on fait vieillir¹. Ainsi :

$$\begin{aligned} x_j(t_2) &= \eta_1 e^{-\mu_S(t_2-t_1)} e^{i \arctan \mu_T(t_2-t_1)} + \eta_2 \\ &= \rho e^{i\phi} \end{aligned}$$

et plus tard :

$$x_j(t) = \rho e^{-\mu_S \tau_2} e^{i \arctan(\tan(\phi) + \mu_T \tau_2)}$$

$$i = \sqrt{-1}, \tau_2 = t - t_2 \text{ et } \mu_S = \mu_T = 1/TW$$

Si de nouvelles impulsions arrivent, on recommence la même opération.

Une fois le vecteur d'entrée X calculé, le potentiel est égal soit au produit scalaire hermitien :

$$V(X, W) = \sum_{j=1}^N \overline{w_j} \cdot x_j$$

avec $\overline{w_j}$ le complexe conjugué de w_j

soit à la distance hermitienne :

$$D(X, W) = \sqrt{\sum_{j=1}^N (x_j - w_j) \overline{(x_j - w_j)}}$$

Ces fonctions potentielles dans les complexes sont à associer respectivement au produit scalaire et à la distance euclidienne pour les modèles classique dans les réels.

La fonction d'activation F appliquée au potentiel détermine alors la sortie y . Lorsque la distance hermitienne est utilisée comme fonction potentielle, le résultat est réel et les fonctions de transfert classiques sont utilisées. Quand c'est le produit scalaire hermitien, le résultat étant en général complexe, des fonctions particulières, respectant certaines propriétés adaptées au corps des complexes [4], doivent être utilisées. La fonction retenue a été proposée dans [9]

$$F(x + iy) = p \cdot x + ip \cdot y$$

$$p = \frac{\tanh(\sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}}$$

Cette fonction, qui applique une tangente hyperbolique au module sans changer la phase du

¹ce qui est différent de "faire vieillir les impulsions indépendamment puis de les sommer" à cause de la fonction arctan (voir [1])

complexe, permet un bon compromis entre les différents critères nécessaires pour assurer un bon apprentissage une fois ces neurones intégrés dans des architectures neuronales comme le perceptron multicouche (voir ci-dessous).

1.2 Intégration dans des réseaux : les STANN

Disposant toujours d'une algèbre avec la notion de produit scalaire et de distance, il a été possible d'intégrer ce modèle de neurones dans des architectures courantes de réseaux de neurones, en adaptant assez facilement les algorithmes d'apprentissage et d'exploitation à l'algèbre complexe. Ainsi, [10] et [1] présentent des versions spatio-temporelles pour le perceptron multicouche (ST-MLP), pour les réseaux à bases radiales (ST-RBF), pour les réseaux de Reilly, Cooper et Elbaum (ST-RCE), pour les cartes auto-organisées de Kohonen (ST-Kohonen) et pour leurs versions sans voisinage (ST-Kmeans).

Une procédure générale d'utilisation de ces réseaux utilisant la distance hermitienne pour la classification de signaux spatio-temporels a ensuite été proposée [1]. Le système est présenté Figure 2.

Le module de prétraitement a pour but de générer des séquences d'impulsions sur la base des signaux d'entrée. Le module de Quantification Vectorielle émet des impulsions lorsqu'il identifie des sous-séquences (d'une taille $TW1$ unités temporelles) qu'il aura appris à reconnaître (à l'aide d'un ST-Kmeans ou ST-Kohonen). Le dernier module se charge de la classification et génère des impulsions caractéristiques de la classe des séquences d'impulsions (d'une taille $TW0$ unités temporelles) qu'on lui aura appris à identifier (avec un ST-RCE ou ST-MLP).

Le frein à l'utilisation de cette procédure et de ces outils était le module de prétraitement pour lequel aucune solution générique n'était présentée et qui nécessitait donc encore un gros travail d'analyse et de validation.

2 La génération d'impulsions

2.1 Problématique

Comme nous l'avons vu, les STANN font partie des récents modèles de réseaux de neurones prévus pour accepter en entrée des événements discrets, c'est à dire des impulsions (spikes, voir Figure 3-d). Parmi les autres modèles utilisant des telles entrées,

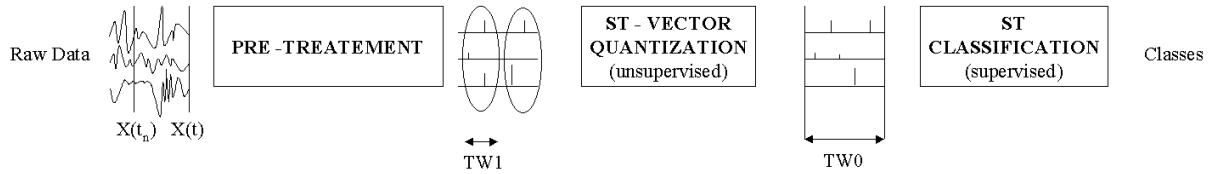


FIG. 2 – Système général de classification par des STANN

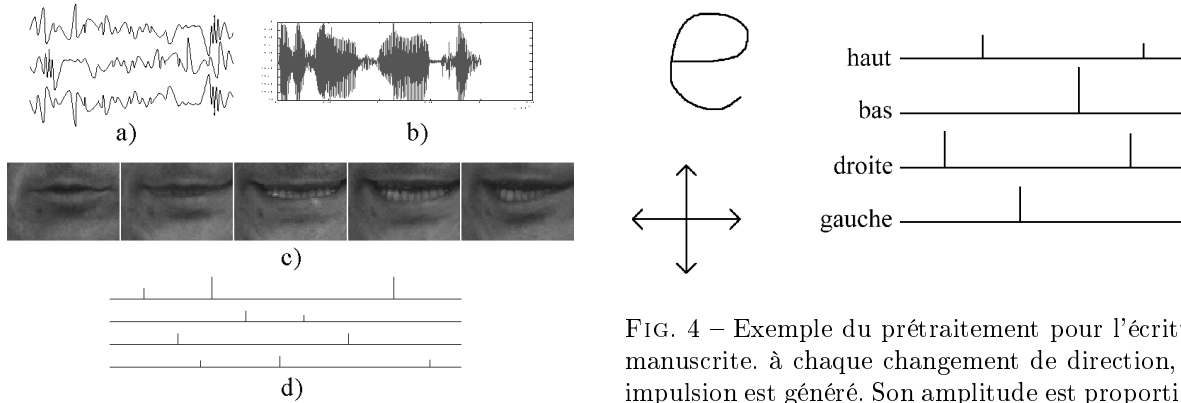


FIG. 3 – Différents signaux : a) signal multicapteurs, b) signal de parole, c) séquence d’image, d) signal impulsionnel

les plus connus sont les I&F (Integrated and Fire, [6]) et les PCNN (Pulses Coupled Neural Network, [5]).

Mais dans la pratique, il existe très peu de signaux bruts impulsionnels. Les capteurs ont en général été conçus pour justement avoir en sortie des signaux continus et non impulsionnels. D’ailleurs pour les rares capteurs générant des impulsions (capteur de déplacement dans les souris d’ordinateur par exemple), les signaux produits sont immédiatement traités par un intégrateur afin de générer un signal en échelon. Quant à une méthode générique de conversion en signaux impulsionnels, il n’y en a aucune.

Une première solution consiste à utiliser des signaux statiques et à les convertir en information temporelle. Citons comme exemple [7] qui fait de la segmentation d’images avec des PCNN. La génération d’impulsions espacées dans le temps se fait en convertissant les niveaux de gris des images en information temporelle.

Lorsqu’il s’agit de traiter des signaux de nature spatio-temporelle (séquences d’images, signaux multicapteurs, etc. voir Figure 3-a-b-c) qui évoluent

FIG. 4 – Exemple du prétraitement pour l’écriture manuscrite. à chaque changement de direction, un impulsion est généré. Son amplitude est proportionnel au temps durant lequel le stylet s’est déplacé dans la direction en question

de façon continue dans le temps (niveau de luminosité d’un pixel pour les images, niveau sonore pour la parole, etc.), il faut appliquer une opération de prétraitement pour convertir ces signaux en signaux impulsionnels. Ainsi, en écriture manuscrite dans [11], une impulsion est générée à chaque changement de direction (les 4 directions haut, bas, droite et gauche ont été retenues, voir Figure 4). L’amplitude de l’impulsion est proportionnelle au temps pendant lequel la direction a été maintenue. Pour la lecture labiale dans [2], un traitement de l’image permet de suivre quatre points significatifs de la bouche (commissures et centre des lèvres, voir Figure 5). Le suivi de ces points permet d’évaluer leurs mouvements. Lors d’un mouvement, une impulsion est générée (sur l’entrée du neurone caractérisant le point et le sens) au moment où le déplacement est le plus rapide. L’impulsion est proportionnelle à la vitesse du point à ce moment là. Pour la détection de mouvement par des I&F dans [12], une impulsion est générée lorsque le contraste du pixel a changé.

Plus généralement, à chaque fois que l’on modifie le signal utilisé ou le problème traité, il faut redéfinir un prétraitement, ce qui représente une perte de temps considérable en développement et en tests.

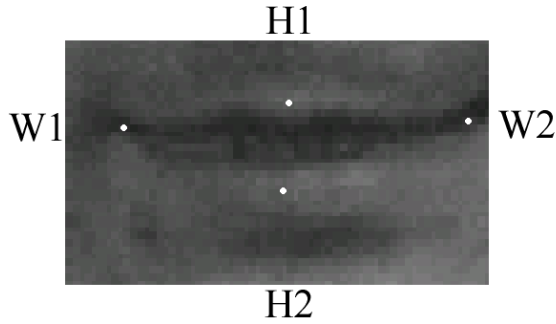


FIG. 5 – Les quatre points suivis sur la bouche pour générer les impulsions

De plus, pour parvenir à générer des impulsions à partir des signaux bruts, on supprime parfois des informations non identifiées comme pertinentes mais dont l'absence va fortement handicaper le système de classification en aval. Pour reprendre l'exemple de la lecture labiale, suivre l'évolution des quatre points sur les lèvres, c'est ne pas pouvoir prendre en compte les asymétries de la bouche, la position des dents et de la langue, ni même le fait qu'elles soient visibles ou pas.

2.2 Une méthode générique de génération d'impulsions

Pour générer simplement des impulsions à partir de signaux évoluant dans le temps de façon continue, nous proposons de réaliser une Quantification Vectorielle (QV) statique sur la base des signaux pris à chaque instant. Cette QV permet d'associer à la forme statique (définie à un instant donné par l'ensemble des capteurs) un prototype de forme. Nous procédons en quatre étapes :

Apprentissage :

- Définition des M prototypes statiques. Un K-means [8] est réalisé sur une base représentative des signaux multicapteurs pris à chaque instant (par exemple : une image dans une séquence est un élément de la base).
- Détermination d'éléments caractéristiques pour chaque prototype P_j . Le plus courant est le calcul d'un rayon d'activité autour du prototype : pour chaque P_j on évalue l'ensemble E_j des exemples de la base d'apprentissage qui lui sont associés, et on détermine le rayon d'activité r_j comme la distance entre P_j et l'exemple la plus éloigné de E_j (comme le fait

[1] pour l'évaluation du rayon dans un STAN utilisant la distance hermitienne).

Exploitation :

- Identification à chaque instant du prototype P_k dont se rapproche le plus le signal statique $X(t)$ en entrée.
- Emission d'une impulsion. Les M sorties du module de prétraitement sont nulles sauf celle qui correspond au prototype k sur laquelle est générée une impulsion. La valeur de cette impulsion dépend d'une "loi de génération d'impulsions" dont les arguments sont l'entrée $X(t)$, le prototype P_k et les éléments caractéristiques associés à P_k . Dans notre cas, l'activation la plus efficace a été la loi gaussienne définie par le rayon d'activité et appliquée à la distance entre l'entrée et le prototype. Ainsi :

$$y_k(t) = e^{-\frac{D^2(X(t), P_j)}{2r_j^2}}$$

En fait, l'utilisateur n'a plus qu'à définir deux paramètres : d'abord le nombre de prototypes à conserver dans la phase de quantification vectorielle ; ensuite la loi de génération d'impulsion, qui va elle-même déterminer les éléments caractéristiques. A notre avis, cette loi de génération d'impulsions dépend principalement du modèle et pas de la nature du signal ou bien de l'application. Par exemple, pour un I&F qui n'a qu'un degré de liberté, il ne faut pas moduler l'impulsion mais la laisser unitaire. Pour un modèle encore plus proche de l'observation biologique qui serait capable de travailler sur la fréquence, une loi générant des impulsions avec une fréquence d'autant plus importante que l'entrée est proche du prototype sera sûrement préférable.

Autre avantage important de la méthode en plus de sa simplicité, c'est que le problème de la robustesse du système est alors ramené au problème de la robustesse de la quantification vectorielle. Ainsi il faut que deux formes identiques soit quantifiées identiquement. Problème non trivial, la robustesse à la quantification vectorielle est néanmoins un problème très étudié, très publié, plus facile à appréhender que le problème de la robustesse d'un prétraitement de génération d'impulsions complet, et une fois une solution envisagée, elle est beaucoup plus facile à tester et à valider. En cela, la méthode est également intéressante car si du travail est encore nécessaire, il est plus rapide et d'un abord plus simple pour un utilisateur voulant utiliser les

STANN comme une boîte noire.

3 Illustration en lecture labiale

3.1 Présentation

En reconnaissance vocale, la modalité audio ne suffit pas toujours lorsque le milieu ambiant est bruyé. Il est alors intéressant de renforcer le système par un module utilisant la modalité vidéo afin de lire sur les lèvres, ce qui améliore la robustesse du système [3]. Les outils les plus souvent utilisés restent les modèles de Markov cachés (HMM : Hidden Markov Model) [17] et les réseaux de neurones statiques où la modalité temporelle est ajoutée en répliquant les entrées avec différents retards (TDNN : Time Delay Neural Network) [13].

Pour éprouver le STANN et leur méthodologie d'application (§1), la lecture labiale a également été étudiée dans l'équipe ([1] et [2]). Le problème a été posé dans un cadre monocuteur, sur la reconnaissance de chiffres prononcés en français. Les conditions d'étude et de test furent les suivantes. Dans un premier temps, des images de toute la partie inférieure du visage sont saisies à 25 Hz via une carte vidéo MJPEG dont le rapport de compression était réglé à 6:1. Elles sont ensuite récupérées et sauvées au format Bitmap 24 bits (RVB, 3x8). Leur taille était de 360x540 pixels. Arrive alors le prétraitement pour générer des impulsions. Comme nous l'avons déjà dit, ce prétraitement est basé sur la détection du mouvement de quatre points (les centres et les commissures des lèvres). Au moment où le déplacement est le plus rapide, une impulsion est générée. Les images étaient désentrelacées et l'étude était restreinte à une portion de l'image (288x192) (l'utilisateur ne devait pas trop bouger la tête). Ensuite, les systèmes de quantification vectorielle spatio-temporelle et de classification spatio-temporelle sont normalement utilisés.

En tout, vingt-six séquences ont été enregistrées en trois sessions à quinze jours d'intervalle (afin d'assurer et vérifier la robustesse aux conditions environnementales). Durant chaque séquence, l'utilisateur prononce les chiffres de 0 à 9 en français. Les séquences numérotées impaires constituent la base d'apprentissage, les séquences numérotées paires constituent la base de test. Chaque séquence est découpée en dix sous-séquences (une pour chaque chiffre). Ainsi, la base d'apprentissage comprend 260 petites séquences (26 pour chaque chiffre) et

la base de test 250 (25 pour chaque chiffre) (voir Figure 6).

L'objectif après apprentissage est que le système reconnaisse correctement le chiffre prononcé durant une sous-séquence de la base de test.

3.2 Applications de notre prétraitement et Résultats

Puisque nous possédions au sein de l'équipe une application, son protocole et ses résultats dans le cadre d'un traitement de signaux spatio-temporels continus par des STANN, nous avons décidé de la reprendre en conservant tout sauf le prétraitement de génération d'impulsion ou nous utilisons notre algorithme générique.

Tout d'abord, il a fallu rendre les données robustes à la quantification vectorielle. Pour cela, nous avons du traquer la bouche pour être insensible au déplacement du visage par rapport à la caméra, selon une méthode courante de poursuite [14]. Ensuite nous avons fait une égalisation d'histogramme afin de ne pas être gênés par les changements de luminosité et de contraste entre les différentes sessions. Notons au passage que nous avons pu nous contenter de travailler sur les images converties en 256 niveaux de gris et sous-échantillonnées d'un rapport huit, ce qui nous faisait à la fin de simples images 36x24x8 bits. Cette taille, trop petite pour le suivi de points ou la détection de contours, est suffisante lorsque l'on utilise la quantification vectorielle.

Ensuite, pour la génération des impulsions, nous avons conservé 20 prototypes pour la quantification vectorielle.

Le reste du système a été conservé, y compris les plages de variation des paramètres. L'ensemble est présenté sur la Figure 7. Le Kmeans étant sensible à l'initialisation qui est aléatoire, les résultats varient entre deux apprentissages successifs, même à nombre de prototypes identiques. Nous avons donc pour chaque réglage fait dix simulations et regardé les résultats en moyenne.

Dans [1], lorsqu'un ST-RCE est utilisé comme classifieur ST, sans exploiter le module de QV-ST (voir Figure 2), les résultats sont de l'ordre de 58% de bonne classification. Dès que le module de QV est utilisé, des paramètres optimaux donne 72% de bonne reconnaissance (77% dans le meilleur des cas). Avec le nouveau prétraitement, l'utilisation du QV n'est plus significative, la classification correcte étant de 85% dans les deux cas (90% dans le

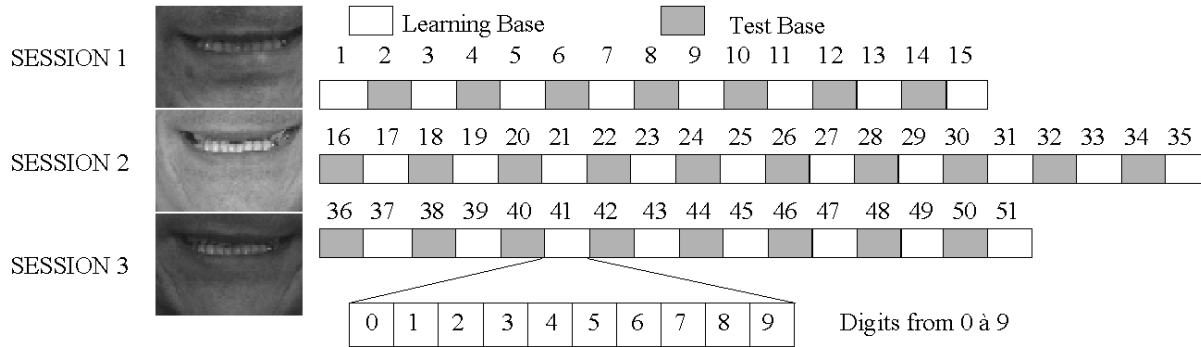


FIG. 6 – La base réalisée sous trois sessions différentes

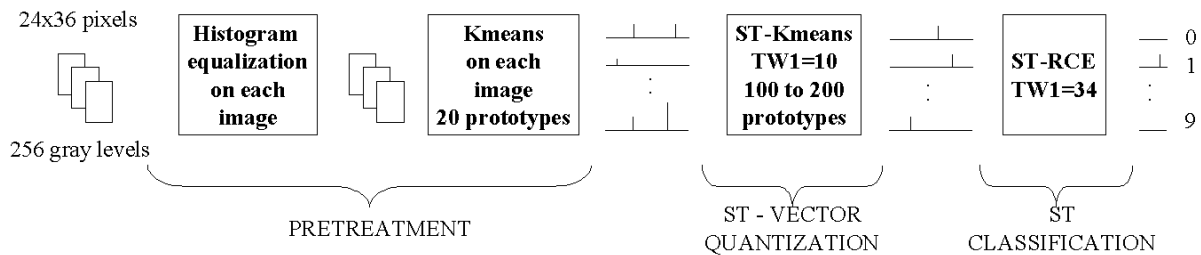


FIG. 7 – La base réalisée sous trois sessions différentes

meilleur des cas avec le module QV-ST).

Au final, non seulement ce prétraitement générique appliqué à la lecture labiale n'a pas handicapé l'ensemble du système, mais les performances ont même été augmentées de 13 points. Et le taux exceptionnel de 85% de bonne reconnaissance en moyenne - même s'il est à prendre avec prudence puisque notre séquence n'a fait l'objet d'aucun benchmark - est très satisfaisant et prometteur quant à l'utilisation du système général de classification par les STANN avec comme prétraitement l'algorithme générique présenté. C'est d'autant plus vrai que l'étude exhaustive des erreurs permet d'isoler majoritairement des confusions entre les chiffres "cinq", "six" et "sept" qui, quand on veut se contenter de la lecture labiale, sont effectivement des éléments très difficiles à séparer puisque les mouvements sont quasi-identiques.

4 Conclusions et perspectives

Jusqu'à présent, le principal facteur qui limitait l'utilisation des STANN était le prétraitement pour passer de signaux évoluant de façon continue dans le temps à des séquences d'impulsions. Le prétraitement que nous avons proposé dans cet article per-

met de réaliser ce passage de façon automatique. Extrêmement simple à mettre en œuvre et ne nécessitant aucune information a priori, il conserve, au moins dans le cadre de la lecture labiale, suffisamment d'information pertinente pour nous permettre d'utiliser la méthodologie d'utilisation des STANN développée dans [1]. A posteriori, une fois les outils basés sur les STANN et la quantification vectorielle implémentés et validés, le seul travail réel aura été de définir un tout premier prétraitement pour rendre les signaux d'entrées robustes à la quantification vectorielle (ici, la translation de la bouche dans l'image), puis d'utiliser simplement les outils. Cette réussite nous permet d'envisager des perspectives de recherches à plusieurs niveaux.

En ce qui nous concerne, nous travaillons actuellement sur une amélioration du tout premier prétraitement pour augmenter la robustesse à la quantification vectorielle et ainsi permettre un apprentissage "un coup". L'idée est d'extraire suffisamment d'informations sur la première séquence pour ensuite pouvoir classifier correctement toutes les autres séquences. Actuellement, les résultats sont très satisfaisants au sein d'une session (apprentissage séquence 1, test sur les 14 suivantes) mais chutent dès que l'on change de session.

Dans le cadre plus général de l'application des

STANN aux interactions homme-machine [16], plusieurs autres études sont à mener. Parmi les plus prometteuses citons une nouvelle étude sur l'écriture manuscrite et la reconnaissance de gestes de la main (proche du langage des sourd-muet). En effet, la première étude sur l'écriture manuscrite avait défini comme prétraitement la quantification de la direction de déplacement du stylet par les directions privilégiées haut, bas, droite et gauche. A posteriori, ce concept est très proche de celui de la quantification vectorielle. En plus, au lieu de directions prédéterminées (haut, bas, droite et gauche), l'algorithme devrait trouver les directions privilégiées (dans le cas d'une écriture penchée) ce qui est susceptible de donner des changements de direction plus significatifs. Nous sommes donc optimistes quant au gain que ce prétraitement pourrait apporter à cette application. Pour la reconnaissance de gestes de la main, sa similarité dans sa forme et ses objectifs avec la lecture labiale (analyse des déformations d'un objet dans une séquence d'image pour déterminer un mot "prononcé"), et la complète indépendance entre le traitement général par les STANN et le problème considéré (pas de prétraitement utilisant le fait qu'il s'agisse d'une bouche ou bien d'une main) nous laisse espérer encore une fois de bons résultats (chose inenvisageable quand le prétraitement était le suivi de points sur les lèvres). De plus, la confrontation du système général à d'autres problèmes nous permettrait de vérifier si - comme nous le pensons - la loi de générations d'impulsion dépend seulement du modèle ou bien si la nature des signaux et de l'application ont encore une influence.

Enfin, mais nous sortons du cadre de notre équipe, il serait intéressant de tester ce prétraitement sur d'autres modèles de neurones à impulsions, en adaptant la loi de génération d'impulsions au modèle. Pour un I&F, générer une impulsion unitaire (et non pas normée par une gaussienne comme pour le STANN) est susceptible de fonctionner. Pour PCNN, manquant d'expérience avec ces réseaux, il nous est impossible de proposer une loi ayant de grande chance de fonctionner. Dans tous les cas, s'il s'avère pertinent, ce prétraitement adapté aux modèles permettrait de leur ouvrir tout un domaine de nouvelles applications qui leur est fermé aujourd'hui par manque de méthodes de conversions en signaux impulsifs.

Références

- [1] A.R. Baig "Une approche méthodologique de l'utilisation des STAN appliquée à la reconnaissance visuelle de la parole" *Université Rennes I*, PhD Report, 2000.
- [2] A.R. Baig, R. Séguier et G. Vaucher "A Spatio-temporal Neural Network applied to visual speech recognition" *ICANN*, 1999.
- [3] C.C. Chibelushi and F. Deravi and J.S.D. Mason "Audio-Visual Integration in Multimodal Communication" *Proceedings IEEE*, Mai 1998.
- [4] G.M. Georgiou and C. Koutsougeras "Complex domain backpropagation" *IEEE trans. on circuits and systems - II : Analog and digital signal processing*, Mai 1992.
- [5] J.L. Johnson and M.L. Padgett "PCNN Models and Applications" *IEEE trans. on neural networks*, Mai 1999.
- [6] W. Gerstner "Time structure of the activity in neural network models" *in Physics Review*, 1995.
- [7] T. Lindblad and J.M. Kinser "Image Processing using Pulse-Coupled Neural Networks" *Springer-Verlag*, 1998
- [8] L. Lebart, A. Morineau et M. Piron "Statistique exploratoire multidimensionnelle" *Dunod*, 1997
- [9] T. Masters "Signal and image processing with neural networks" *John Wiley & Sons, Inc.*, 1994.
- [10] N. Mozayyani "Introduction d'un codage spatio-temporel dans les architectures classiques de réseaux de neurones artificiels - Application à la reconnaissance de caractères manuscrits" *Université Rennes I*, PhD Report, 1998.
- [11] N. Mozayyani, A.R. Baig et G. Vaucher "A Fully Neural Solution for on-Line Handwritten Character Recognition" *IJCNN*, 1998.
- [12] W. Paquier, A. Delorme, R. Vanrullen et S. Thorpe "Detection de mouvements apparents par codage asynchrone" *NSI - Neuroscience et Science de l'Ingénieur*, 2000.
- [13] R. Stiefelhagen and U. Meier and J. Yang "Real-Time Lip-Tracking for Lipreading" *Proc. of Eurospeech*, 1997.
- [14] N.P. Topiwala "Wavelet image and video compression" *KLUWER ACADEMIC*, 1998.

- [15] G. Vaucher “A la recherche d’une algèbre neuronale spatio-temporelle” *Université Rennes I*, PhD Report, 1996.
- [16] G. Vaucher “Une famille de neurones à impulsions appliquée aux interactions homme-machine” *NSI - Neurosciences et Science de l’Ingénieur*, 2000.
- [17] T. Wark and S. Sridharan and V. Chandran “The use of temporal speech and lip information for multi-modal speaker identification via multi-stream HMM’s” *ICASSP*, 2000.

Annexe : résultats de la génération d’impulsions

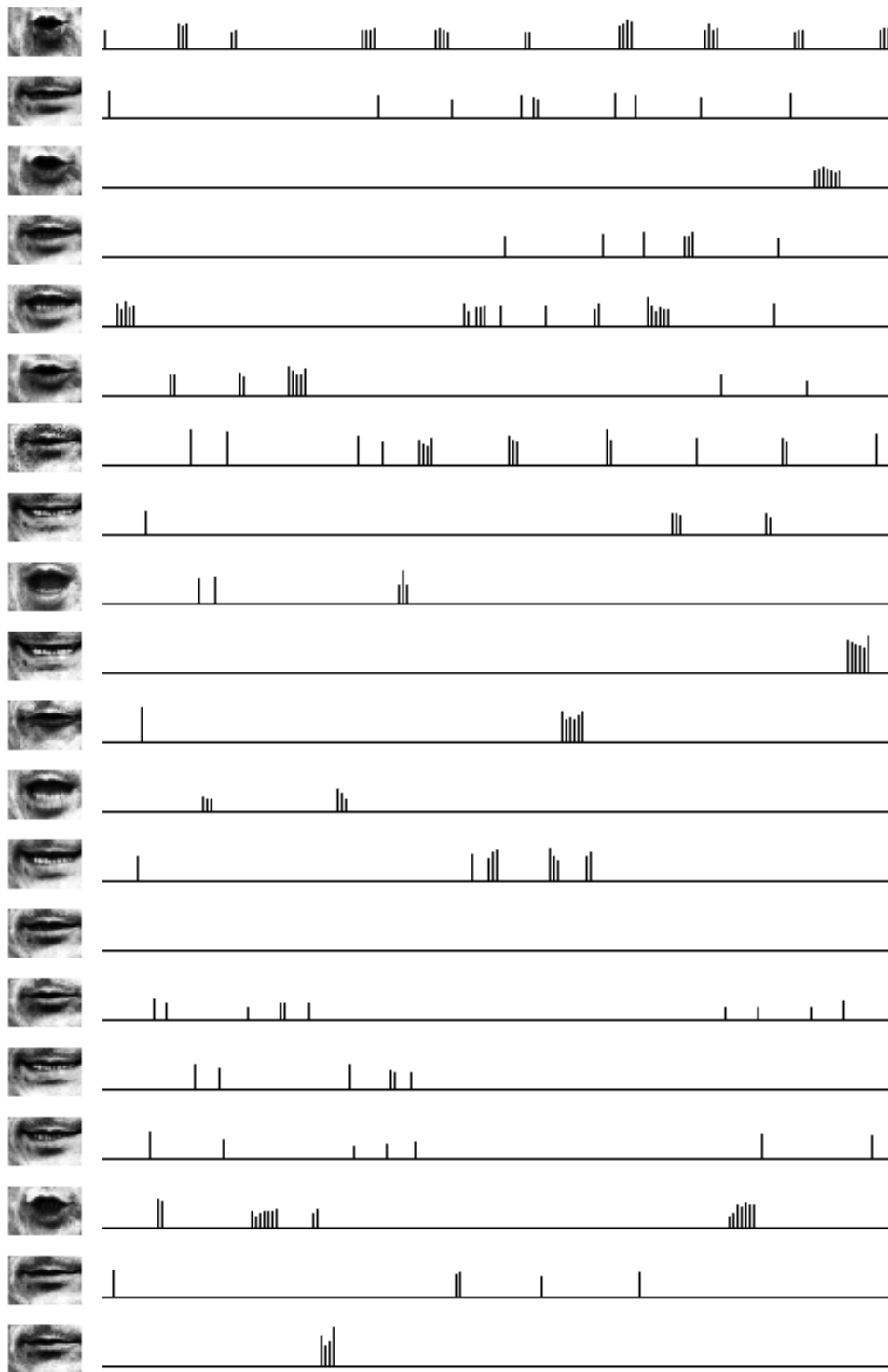


FIG. 8 – Les prototypes obtenus durant l’une des simulations et les impulsions générées pour chaque prototype lorsque l’on présente une séquence test où les chiffres de 0 à 9 sont prononcés dans l’ordre. Le temps s’écoule de gauche à droite. Une et une seule impulsion est générée à chaque instant.

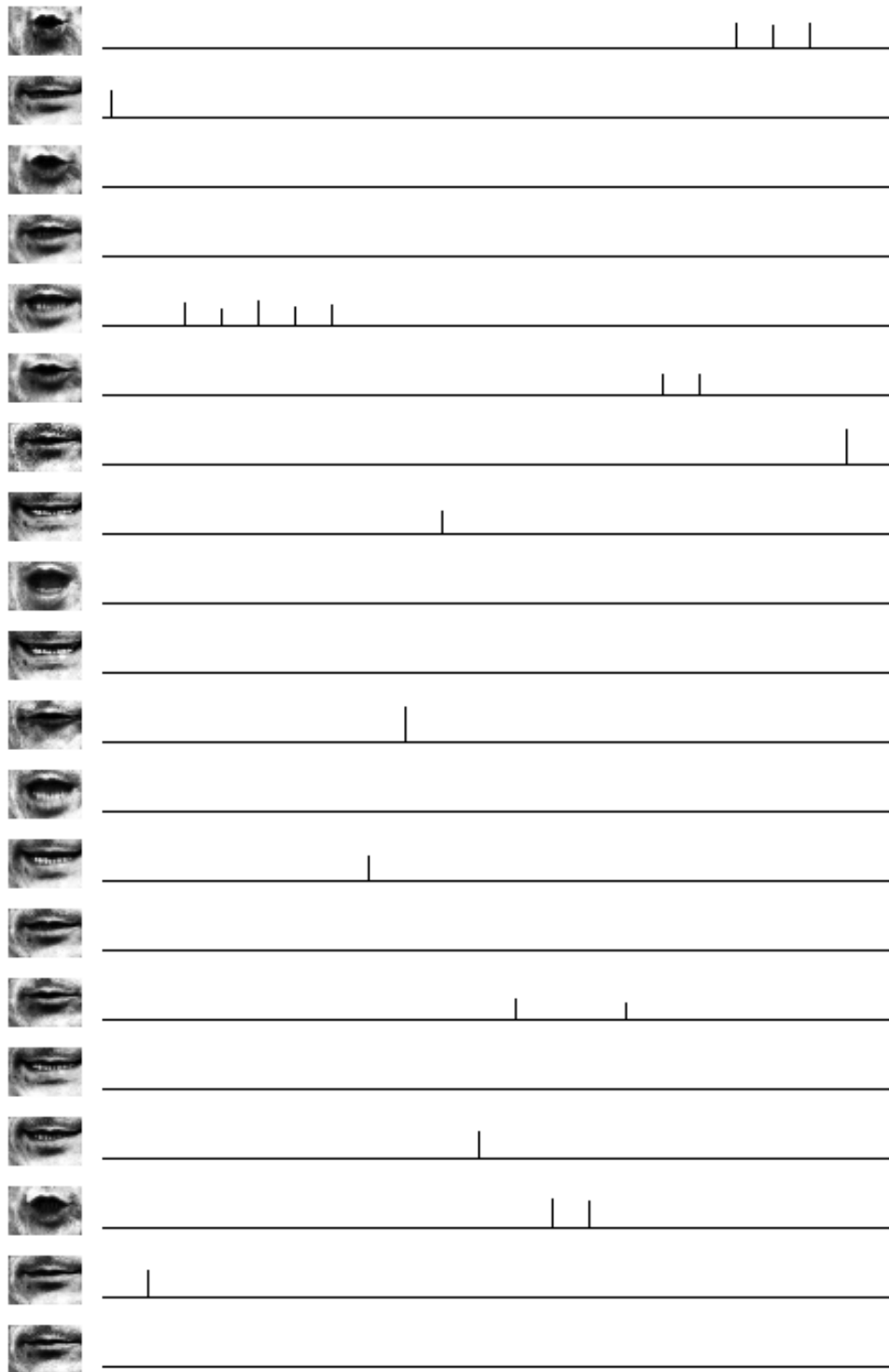


FIG. 9 – Les mêmes prototypes et les impulsions générées pour chaque prototype lorsque l'on présente simplement la sous-séquence correspondant au zéros.