

Classification automatique des voix Parkinsoniennes de patients algériens par réseaux de neurones artificiels

Kamel FERRAT & Ghania GUERTI
École Nationale Polytechnique (ENP) - Alger
CRSTDLA

Introduction

Cet article traite de la classification automatique des voix Parkinsoniennes (VP) de patients algériens par Réseaux de Neurones Artificiels (RNA). La classification automatique des voix pathologiques a reçu une grande attention des chercheurs durant cette dernière décennie (Ritchings, Mc Gillion, & Moore, 2002). Dans les approches récentes de classification de la voix pathologique, différentes méthodes de classification de formes (Pattern Classification) ont été utilisées (Ananthakrishna, Shama & Niranjana, 2004; Godino-Llorente & Gomez-Vild, 2004; Umapathy, Krishnan, Parsa, & Jamieson, 2005; Wang & Jo, 2006). Dans notre étude, nous avons appliqué les RNA pour la discrimination entre les voix normales (VN) et les voix parkinsoniennes (VP). L'idée principale des RNA est de simuler l'organisation des neurones biologiques humains et leurs interconnexions pour traiter les informations (Dreyfus, 2004). Parce que les RNA sont bien adaptés à la reconnaissance des formes, et parce que la reconnaissance vocale est essentiellement un problème de reconnaissance de formes, de nombreux chercheurs ont essayé d'appliquer les réseaux de neurones pour la reconnaissance vocal et par suite donc à la reconnaissance des voix pathologiques.

1. La maladie de Parkinson

1.1 Définition et description

La maladie de Parkinson (MP) est une maladie neurodégénérative (MND) chronique et progressive du système nerveux central, décrite pour la première fois par James Parkinson en 1817. On observe une dégénérescence progressive des neurones d'une petite partie du cerveau, le Locus Niger. Ces neurones ont la fonction de déclencher l'initiation des mouvements moteurs volontaires et d'assurer le bon fonctionnement des mouvements automatiques comme la respiration ou la marche. Il en résulte un ensemble de troubles de motricité tels que des tremblements au repos, une rigidité, une lenteur des mouvements et une perte de réflexes posturaux (Jankovic, 2008), qui se traduisent par des troubles de la voix et de la parole. Cette maladie affecte principalement les personnes âgées de plus de 50 ans, même si au moins 10% des cas surviennent à un âge précoce (30 à 50 ans). En outre, elle affecte environ 6 millions de personnes dans le monde. Elle est la deuxième maladie neurodégénérative la plus répandue dans le monde, après la maladie d'Alzheimer. Une étude de Logeman et al. montre que près de 45% des patients présentent des troubles du langage (dysarthrie), environ 89% présentent des troubles de la voix (dysphonie) et environ 20% présentent une déficience du flux de la parole continue (pauses fréquentes et coupures de mots) (Logeman, Fisher & Boshes, 1978). En outre, certaines études observent une diminution de la capacité des processus cognitifs impliquant la prise de décision comme un symptôme secondaire du syndrome de Parkinson (Lachmann et al., 2008; Mitchell & Bouças, 2009).

1.2 Troubles de la voix et de la parole chez le patient parkinsonien

La difficulté à parler (dysarthrie) est l'un des symptômes les plus courants qui se développent chez les patients atteints de la MP. On estime que près de 89% des patients ont des difficultés avec leur parole, incluant des troubles des fonctions laryngale, respiratoire et articulatoire (Ramig, Fox &

Sapir, 2008). En outre, les caractéristiques de la MP étaient non seulement une plus forte accélération de la vitesse d'articulation dans le cadre de la parole, mais aussi de longues pauses avant les débuts de parole, des silences inappropriés entre les mots, des répétitions non contrôlées de sons, et un faible rythme de la parole (Skoda & Schlegel, 2008). Les voix des patients sont caractérisées par une respiration limitée et un trouble de mobilité des cordes vocales impliquant un pitch hachuré comme le montre la figure 1b, un volume de parole réduit, et une voix monotone (Holmes, Oates, Phyland, & Hughes, 2000).

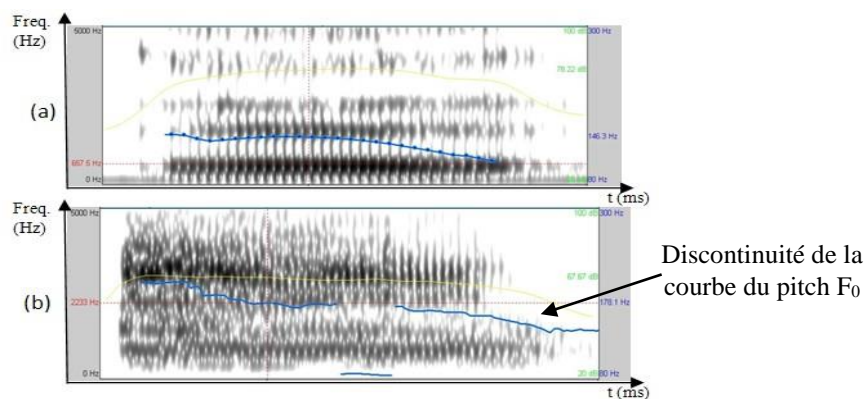


Figure 1. La voyelle [a] en voix normale (a) et en voix parkinsonienne (b).

En outre, nous remarquons un niveau faible d'intensité moyenne, en particulier pour les patients atteints d'une forte dysarthrie (Sapir, Ramig, & Fox, 2006) et enfin une diminution significative du Temps maximum de Phonation (TMP) en comparaison avec les données normatives (Locco, 2005). La consonne occlusive est réalisée de manière incomplète par les articulateurs, en raison d'un défaut de la fermeture du conduit vocal. Cette occlusion partielle donne lieu à des fuites d'air, source de bruit de friction lors de la réalisation de l'occlusive. En outre, l'économie du geste articulaire provoque un changement dans le lieu et/ou mode d'articulation (Rigaldie, Nespoulous, et Vigouroux, 2004). Pour aider le patient à parler correctement, nous utilisons des techniques spécialisées de rééducation dont la plus connue est la technique LSVT (Lee Silverman Voice treatment) (Kreiman & Gerratt, 2005). Pendant les séances de rééducation, les exercices sont effectués de façon répétée. Ces exercices tiennent compte notamment du temps maximum de phonation, de la production de sons les plus graves et les plus aigus, et la répétition à voix élevée de phrases.

2. Application des RNA à la reconnaissance et à la classification des voix pathologiques

Dans ce présent article, nous ne décrivons pas les réseaux de neurones artificiels RNA, une étude détaillée a été déjà présentée lors du 1^{er} Congrès sur les neurosciences (Ferrat & Guerti, 2011). Dans le cadre de cette présente étude, un réseau de neurones à délais temporels TDNN (Time Delay Neural Network) a été utilisé pour reconnaître et classer un ensemble de données mélangées comprenant des voix normales VN et des voix parkinsoniennes VP. Cette méthode dite TDNN nous permet la discrimination de la voix en prenant en compte de l'aspect dynamique de la parole et par conséquent du phénomène de la coarticulation. Au cours de la phase dite « d'apprentissage », nous avons utilisé la technique de retropropagation de l'erreur par régularisation bayésienne (BR) et basée sur l'algorithme de Levenberg-Marquardt (LM). Pour les tests de reconnaissance, le vecteur d'entrée est classé avec le vecteur associé à la classe qui donne une distance totale minimale.

2.1 Participants

Trois locuteurs normaux et cinq locuteurs parkinsoniens ont participé aux expériences acoustiques. Quatre patients parkinsoniens n'avaient pas bénéficié d'une rééducation par un orthophoniste. Un seul patient a bénéficié d'une rééducation (sessions pendant 2 ans). En outre, tous les patients présentent des troubles du langage à divers degrés, et utilisent le dialecte arabe comme moyen de communication (tableau 1). Nous aurons à comparer ces cas pathologiques au groupe de référence ne présentant aucun trouble de la voix et de la parole.

| Caractéristiques des patients | Age | Rééducation Oui(+)/Non(-) | Début de la maladie |
|-------------------------------|-----|------------------------------|---------------------|
| 1 | 82 | - | 2005 |
| 2 | 74 | - | 2004 |
| 3 | 68 | - | 2003 |
| 4 | 68 | - | 2001 |
| 5 | 61 | + (2 ans) | 1992 |

Tableau 1. Présentation des patients parkinsoniens

2.2 Analyse acoustique de la VP

La voix a connu un grand intérêt aussi bien dans l'étude de sa physiologie, de sa physiopathologie mais surtout de ses moyens d'exploration et de mesure. Chaque mesure apporte des informations différentes sur les aspects de la production sonore. Une grande complexité des phénomènes physio-acoustiques et des phénomènes psycho-acoustiques de la production vocale rendent impossible l'élaboration d'une méthode unique de mesure et de quantification de la voix. En conséquence, aucune mesure ne suffit à elle seule à caractériser la voix pathologique.

Acoustiquement, la parole normale est analysée essentiellement par l'observation des principaux paramètres suivants: la fréquence fondamentale F_0 , l'intensité et les formants. Cependant l'analyse d'une voix pathologique fait appel, en addition, à d'autres paramètres aussi importants basés sur la technique d'analyse cepstrale du signal vocal. Parmi ces paramètres, nous avons utilisé le degré de perturbation de F_0 (Jitter) et le degré de perturbations de l'intensité (Shimmer) (Klingholz, & Martin, 1985). Ces paramètres acoustiques ont été mis au point au laboratoire en utilisant le logiciel de programmation Matlab[®]. La méthode pour déterminer les paramètres Jitter et Shimmer est principalement basée sur la détection de pics glottiques au moyen de la technique dite "d'analyse cepstrale". L'analyse de F_0 (Hz), du Jitter (%), du Shimmer (% et dB) et de l'intensité (dB) a été réalisée à partir de productions isolées (syllabes et mots) et de voyelles soutenues ([a], [i], [u]).

Chaque voyelle a été produite 5 fois et enregistrée au moyen du sonographe CSL4300B.

2.2.1 Perturbation de F_0 (Jitter)

Le Jitter est une mesure objective qui rend compte de la déviation moyenne relative du pitch F_0 . Ce paramètre représente un meilleur moyen d'explorer la stabilité de la fréquence fondamentale. Le *Jitter* réfère à la variation cycle par cycle de F_0 , dans une trame du signal. Dans notre analyse, le Jitter se calcule par la moyenne de la différence de F_0 entre deux cycles de vibrations consécutifs.

$$J = \frac{1}{N-1} \sum_i |P(i) - P(i + 1)| \quad (1)$$

Le Jitter factor est un meilleur indice pour explorer la stabilité de la fréquence fondamentale. Il permet de normaliser le Jitter moyen en le comparant à la F_0 moyenne.

$$J_f = 100. \frac{J}{\frac{1}{N} \sum_i P(i)} \quad (\%) \quad (2)$$

Avec $P_{moy} = \frac{1}{N} \sum_i P(i)$ alors :

$$J_f = 100. \frac{J}{P_{moy}} \quad (\%) \quad (3)$$

Avec respectivement J : Jitter; N: Nombre de Périodes; P: Période; J_f : Jitter Factor.

Le Jitter augmente, en principe, en cas de pathologie vocale, et une valeur supérieure à 1.04% correspond à une voix pathologique (valeur-seuil mentionnée dans le Manuel du logiciel Praat). En d'autres termes, le seuil normal/pathologie correspond à 1.04%.

2.2.2 Perturbation de l'intensité (Shimmer)

Le *Shimmer* réfère à la variation cycle par cycle de l'intensité, dans une trame du signal. Dans notre analyse, le Shimmer se calcule par la moyenne de la différence d'amplitude entre deux cycles de vibrations consécutifs.

$$S = \frac{1}{N-1} \sum_i |A(i) - A(i+1)| \quad (4)$$

Comme pour le Jitter factor, le Shimmer factor est un meilleur indice pour explorer la stabilité de l'intensité. Il permet de normaliser le Shimmer moyen en le comparant à l'amplitude moyenne.

$$S_f = 100. \frac{S}{\frac{1}{N} \sum_i A(i)} \quad (\%) \quad (5)$$

Avec $A_{moy} = \frac{1}{N} \sum_i A(i)$ alors :

$$S_f = 100. \frac{S}{A_{moy}} \quad (\%) \quad (6)$$

Le Shimmer absolu moyen, exprimé en dB, est la moyenne des rapports d'amplitudes entre deux cycles de vibrations consécutifs. De même que pour le Jitter, le Shimmer factor relativise le Shimmer moyen et augmente en cas d'anomalie laryngée :

$$S_{dB} = \frac{1}{N-1} \sum_i |20 \log_{10}(A(i)) - 20 \log_{10}(A(i+1))| \quad (7)$$

$$S_{dB} = \frac{1}{N-1} \sum_i |20 \log_{10}(A(i)/A(i+1))| \quad (\text{dB}) \quad (8)$$

$$S_{f_{dB}} = 100. \frac{S}{\frac{1}{N} \sum_i 20 \log_{10}(A(i))} \quad (\text{dB}) \quad (9)$$

Avec $A_{dB_{moy}} = \frac{1}{N} \sum_i 20 \log_{10}(A(i))$ alors :

$$S_{f_{dB}} = 100. \frac{S_{db}}{A_{db_{moy}}} \quad (\text{dB}) \quad (10)$$

Avec respectivement S: Shimmer; N: Nombre de pics glottiques; A(i): Amplitude du pic P_i ; S_f : Shimmer Factor.

Comme pour le Jitter, le Shimmer augmente, en principe, en cas de pathologie vocale, et une valeur supérieure à 3.81% correspond à une voix pathologique (valeur-seuil mentionnée dans le Manuel de Praat). En décibel, ce seuil correspond à 0.35 dB.

2.2.3 Architecture de notre système de classification par RNA

Dans ce travail, nous avons utilisé le réseau temporel à délais (TDNN) comme RNA. Cette architecture a été introduite pour la première fois par A. Waibel pour la reconnaissance de la parole (Waibel et al., 1989). Cette méthode suppose que pour la modélisation des signaux dynamiques tels que la parole, il est nécessaire d'introduire de la mémoire dans le réseau. Ces réseaux sont donc capables de traiter des séquences de parole avec une introduction de retards temporels fixes sur les intrants. L'idée principale de ces réseaux est d'apprendre la structure temporelle des événements acoustiques et les relations entre ces événements (comme par exemple les phénomènes de coarticulation dans le cas de la parole).

Pour la phase d'apprentissage, nous avons utilisé la technique de retropropagation de l'erreur (apprentissage supervisé) par Régularisation Bayésienne (BR), exploitant l'algorithme de Levenberg-Marquardt (LM). La retropropagation BR permet de surmonter les problèmes de sous et sur apprentissage du système. L'algorithme LM donne de meilleures performances par rapport aux autres algorithmes d'apprentissage (Fun et Hagan, 1996). Enfin, une classification permet la discrimination entre les VN et les VP, comme le montre la figure 2. Le réseau de neurones, l'algorithme d'apprentissage et tous les autres programmes (les distances, Préaccentuation, la détection des frontières des mots, le Jitter, le Shimmer, l'énergie etc.) sont mis en œuvre à partir du logiciel de programmation Matlab[®].

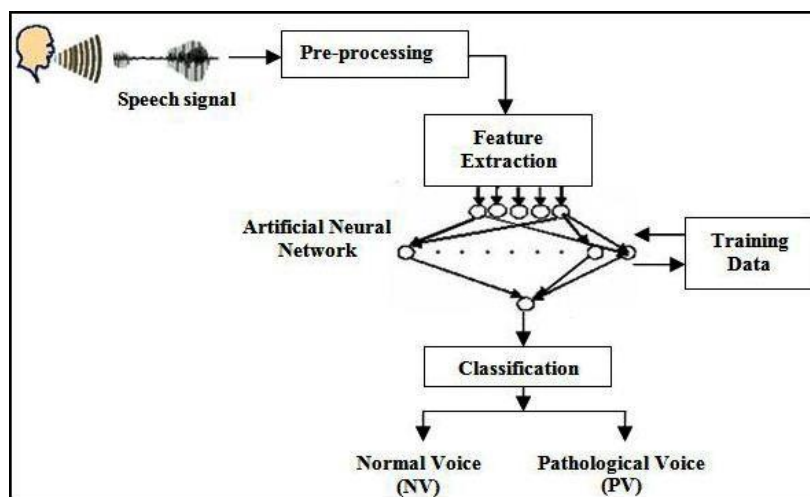


Figure 2. Algorithme de discrimination entre les voix normales (VN) les voix parkinsoniennes (VP).

2.2.4 Base de Données Sonores

Nous avons exploité un corpus de 160 fichiers sonores prononcés par cinq locuteurs parkinsoniens et trois locuteurs normaux ne présentant aucun trouble pathologique. Les enregistrements ont été réalisés au Laboratoire de phonétique et traitement automatique de la parole, du CRSTDLA (Alger). Ces fichiers sont répartis équitablement sur les deux phases d'apprentissage et de reconnaissance. La base de données sonore exploitée lors de la phase d'apprentissage se compose de 40 fichiers sonores pour les VN et 40 fichiers sonores pour les VP. Afin d'évaluer les performances de notre système, nous avons enregistré un ensemble de 80 autres fichiers sonores, avec une fréquence d'échantillonnage de 11025 Hz. Ces fichiers sont différents de ceux utilisés pour la phase d'apprentissage, également répartis équitablement entre les différentes VN et VP. Ces

enregistrements ont été réalisés au laboratoire, dans un environnement naturel contenant du bruit environnant. Nous avons utilisé comme outil d'enregistrement, le sonographe Kay Elemetrics CSL (Computer Speech Lab) 4300B. Nous avons complété quelques enregistrements au niveau de l'hôpital de Beni-Messous d'Alger, en utilisant le logiciel Praat.

Le corpus choisi contient un ensemble de phrases, mots, syllabes et voyelles à l'état isolé, extraits du corpus conçu par N. Zellal (1984) pour l'analyse des VP en orthophonie. Ce corpus tient compte essentiellement des phonèmes spécifiques de l'arabe dans les différents contextes. De plus, l'avantage de ce corpus est conçu en arabe algérois. Ce qui est indiqué dans le cadre de notre travail car les patients sur lesquels nous avons réalisé l'analyse acoustique sont tous arabophones et utilisent l'arabe dialectale algérois.

2.2.5 Prétraitement acoustique et extraction des caractéristiques

Dans une première étape, nous avons appliqué une préaccentuation au signal. Cette phase a pour objectif d'augmenter la quantité d'énergie dans les hautes fréquences et d'avoir une compensation de filtrage des effets de l'acquisition du signal. Pour cela, le signal parole enregistré est appliqué à l'entrée d'un filtre de premier ordre FIR. En outre, en raison de la non-stationnarité du signal parole, il est nécessaire d'extraire les caractéristiques spectrales sur une courte fenêtre (trame de parole) sur laquelle nous supposons que le signal est stationnaire. Pour cette raison, le signal parole est observé à travers une fenêtre d'analyse de longueur finie, appelé fenêtre de Hamming, déplacée régulièrement par pas le long des échantillons de parole. Dans notre cas, la taille de la fenêtre d'analyse et le pas sont respectivement égal à 30 et 10 ms.

Ensuite, une détection automatique des frontières des mots doit être réalisée avant de passer à l'extraction des caractéristiques acoustiques. En d'autres termes, nous devons éliminer tous les échantillons du signal enregistré qui ne sont pas de la parole et de délimiter le signal utile du début à la fin, comme le montre la figure 3. Le défi consiste à éliminer ces échantillons inutiles à partir du signal sans perdre ou fausser l'information pertinente véhiculée par le signal parole. Une fonction procédure utilise un seuil minimal d'énergie moyenne calculée sur la base d'enregistrements de différents bruits d'environnement. Dès que l'énergie dépasse un seuil minimal dans une trame de 30 ms, nous considérons que le début de parole commence à partir de cette trame. Toutes les autres trames précédentes sont éliminées. La même procédure est appliquée à la fin du signal parole.

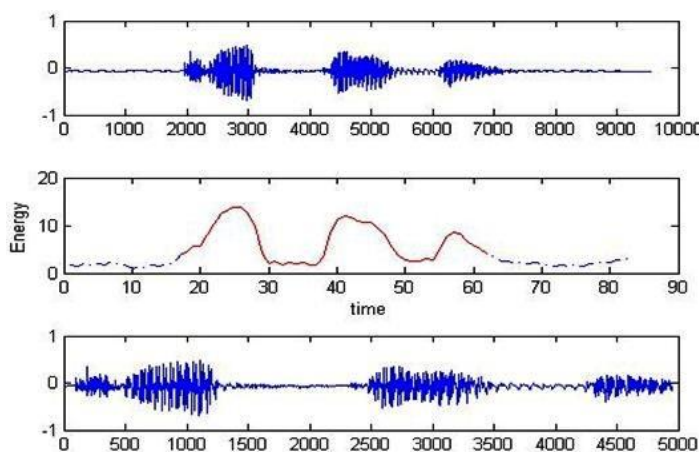


Figure 3. Élimination des trames inutiles de début et fin du mot [kataba]

3. Phase d'apprentissage

Le problème central et crucial dans les RNA est le développement d'algorithmes d'apprentissage et de reconnaissance qui peuvent effectuer l'interaction souhaitée avec un environnement aléatoire ou inconnu et flou, comme le traitement de la parole et la reconnaissance des formes. Lors de la phase d'apprentissage, nous avons appliqué une technique de rétropropagation de l'erreur. Pour atteindre cet objectif, nous propageons un vecteur d'entrée, comme le montre la figure 4, puis nous calculons l'erreur à la sortie eu égard au "vecteur de sortie désiré", et ainsi nous corrigeons à chaque fois les poids afin de minimiser cette erreur. L'apprentissage est arrêté lorsque l'erreur de validation augmente subitement et devient beaucoup plus élevée que la dernière fois où elle a été vérifiée. Une somme carré des erreurs SSE (Sum Squared Error) est ensuite utilisée pour comparer les performances des modèles de réseaux neuronaux (Werbos, 1990). La SSE est définie comme une différence carré entre la sortie désirée (cible) et la sortie réelle (obtenue) du réseau de neurones:

$$SSE = \sum_i (d_j - S_j)^2 \quad (4)$$

avec d_j sortie désirée (sortie cible) pour le neurone j et S_j sortie obtenue (sortie réelle).

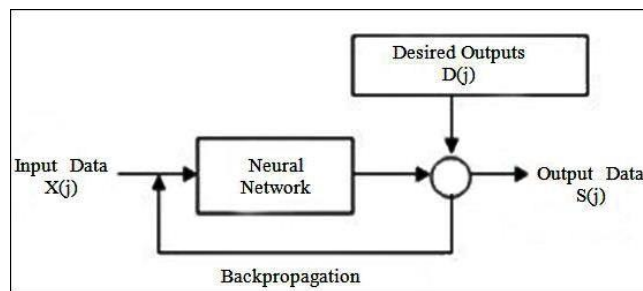


Figure 4. Rétropropagation de l'erreur lors de la phase d'apprentissage.

Pour notre étude, nous avons utilisé l'approche de la technique de rétropropagation BR combinée avec l'algorithme LM, pour minimiser la SSE et pour ajuster les poids W (Fun & Hagan, 1996; Prévoir & Hagan, 1997).

Nous notons que les caractéristiques d'entrée sont normalisées entre $[-1, 1]$. En effet, de meilleures performances de classification sont obtenues en choisissant la valeur moyenne des vecteurs d'entrée du système proche de zéro et de variance moyenne unité (Foresee, Hagan, 1997; Povinelli, Johnson, Lindgren, Ye, 2004). Ces valeurs normalisées sont à proximité de régions de transition d'une fonction sigmoïde, qui permet un plus rapide apprentissage du système.

4. Phase de tests de classification

Pour les tests de classification, nous avons suivi les étapes allant de la lecture des sons enregistrés jusqu'aux tests de reconnaissance, en incluant les différentes étapes de la détection des frontières des mots, de la préaccentuation, et de l'extraction des caractéristiques acoustiques. Pour valider nos tests de classification, nous avons enregistré un ensemble de fichiers sonores contenant des VN et VP, différents de ceux utilisés pour l'apprentissage. Ces enregistrements ont été réalisés au laboratoire avec un milieu naturel contenant du bruit environnant.

Une simulation a été réalisée avec l'ensemble de fichiers tests, afin d'évaluer les performances de notre système lorsque les échantillons inconnus sont présentés à l'entrée. Le comportement de notre classificateur automatique est évalué en termes de pourcentage de classification correcte de l'ensemble de fichiers tests présentés à l'entrée. La méthode pour calculer le taux de reconnaissance (TR) des fichiers tests est donnée par:

$$RA (\%) = (Cas\ Corrects / Total) \times 100 \quad (5)$$

Pour évaluer les performances de notre système, une procédure de classification a été appliquée en changeant à chaque les caractéristiques acoustiques d'entrée, jusqu'à obtenir un meilleur taux de classification. Comme paramètres acoustiques, nous avons utilisé les MFCC (Mel Frequency Cepstral coefficients) connus comme assez représentatifs du signal parole car ils permettent une modélisation du signal parole par des filtres conformes à notre système auditif (Dev & Parmanad 2010; Tiwari 2010), le Jitter, el Jitter factor, le Shimmer, le Shimmer factor et enfin l'énergie.

Nous injectons les paramètres acoustiques choisis à l'entrée du système de reconnaissance, en tenant compte des poids de chaque fichier de parole mémorisé durant la phase d'apprentissage (Ferrat and Guerti, 2011). A chaque étape, nous faisons un apprentissage en utilisant une matrice donnée de vecteurs acoustiques extraits à partir d'un ensemble de 80 fichiers apprentissage (40 fichiers parkinsoniens et 40 fichiers normaux), puis nous passons à la classification automatique, en utilisant d'autres fichiers de paroles inconnus du dictionnaire d'apprentissage, soit 80 autres fichiers répartis équitablement entre les VP et VN. Soit un corpus total de 160 fichiers sonores. Comme vecteurs acoustiques d'entrée, nous avons appliqué respectivement les paramètres MFCC, le Jitter (J), le Jitter factor (J_f), le Shimmer (S), le Shimmer factor (S_f), l'énergie du signal (E) et enfin le taux de passage par zéro (ZC).

| ZC, Jitter Jitter _{factor} , Shimmer, Shimmer _{factor} | Normale | Parkinsonienne |
|-----------------------------------------------------------------------------|-----------|----------------|
| Normale | 29 | 11 |
| Parkinsonienne | 14 | 26 |

Tableau IV.1. Classification Parole Parkinsonienne avec (TPZ, Jitter, Jitter_{factor}, Shimmer, Shimmer_{factor}) vecteurs acoustiques d'entrée

Nous remarquons que sur 40 fichiers sonores correspondant à la voix normale, 29 ont été reconnus correctement alors que 11 fichiers ne sont pas reconnus et considérés comme parole pathologique. De même, sur 40 fichiers de voix parkinsonienne, 26 ont été reconnu et 14 reconnus comme parole normale. Comme ces taux de reconnaissance sont faibles, nous ajoutons à chaque fois, un nouveau paramètre pour améliorer notre système de reconnaissance.

| Energie, Jitter Jitter _{factor} , Shimmer, Shimmer _{factor} | Normale | Parkinsonienne |
|----------------------------------------------------------------------------------|-----------|----------------|
| Normale | 37 | 03 |
| Parkinsonienne | 02 | 38 |

Tableau IV.2. Classification Parole Parkinsonienne avec (Energie, Jitter, Jitter_{factor}, Shimmer, Shimmer_{factor}) vecteurs acoustiques d'entrée

| Energie, ZC, Jitter Jitter _{factor} , Shimmer, Shimmer _{factor} | Normale | Parkinsonienne |
|--------------------------------------------------------------------------------------|-----------|----------------|
| Normale | 34 | 06 |
| Parkinsonienne | 04 | 36 |

Tableau IV.3. Classification Parole Parkinsonienne avec (Energie, TPZ, Jitter, Jitter_{factor}, Shimmer, Shimmer_{factor}) vecteurs acoustiques d'entrée

| MFCC, Energie, ZC, Jitter Jitter _{factor} , Shimmer, Shimmer _{factor} | Normale | Parkinsonienne |
|--------------------------------------------------------------------------------------------|-----------|----------------|
| Normale | 32 | 08 |
| Parkinsonienne | 01 | 39 |

Tableau IV.4. Classification Parole Parkinsonienne avec (MFCC, Energie, TPZ, Jitter, Jitter_{factor}, Shimmer, Shimmer_{factor}) vecteurs acoustiques d'entrée

Finalement, nous avons obtenu les résultats suivants, comme taux de reconnaissance (TR) (%) et taux global de reconnaissance (TGR) (%):

| Paramètres Cas | J, J _f , S, S _f , ZC | J, J _f , S, S _f , E | J, J _f , S, S _f , ZC, E | J, J _f , S, S _f , ZC, E, MFCC |
|--------------------------|-----------------------------------------------|----------------------------------------------|--------------------------------------------------|--------------------------------------------------------|
| Normale (TR %) | 72,50 | 92,50 | 85,00 | 80,00 |
| Parkinsonienne (TR %) | 65,00 | 95,00 | 90,00 | 97,50 |
| TGR (%) | 68,75 | 93,75 | 87,50 | 88,75 |

Tableau 3. Taux de Reconnaissance (TR) et Taux Global de Reconnaissance (TGR). Parole Parkinsonienne avec différents vecteurs acoustiques d'entrée

En se basant sur ces résultats, nous pouvons dire que le système de reconnaissance proposé, à base RNA, donne un bon pourcentage de reconnaissance de VP par rapport aux VN. De plus, en comparant les différents résultats obtenus lorsque nous faisons varier les différents paramètres acoustiques, nous pouvons dire que les meilleurs résultats sont ceux obtenus en utilisant un choix de l'ensemble (E, J, J_f, S et S_f) en tant que caractéristiques acoustiques. Cela montre que ces dernières caractéristiques sont pertinentes dans la discrimination de la parole pathologique par rapport à la parole normale. Ainsi, il est clair que la voix de patients parkinsoniens présente un pitch très irrégulier, une raucité de la voix, un volume de parole réduit et une intensité irrégulière. Ces résultats confirment les résultats des études précédentes rapportées dans la littérature sur la maladie de Parkinson (Holmes et al., 2000; Sapir et al., 2006).

Conclusion

Un schéma de classification entre les voix normales et les voix pathologiques a été présenté dans cette présente étude. Les résultats obtenus nous montrent que les réseaux de neurones artificiels s'adaptent bien à la reconnaissance et à la classification des voix pathologiques. Il

reste que l'un des objectifs essentiels de cette étude est de permettre la caractérisation des voix pathologiques dans domaine acoustique. Le résultat attendu d'une telle étude est de fournir une meilleure compréhension, d'un point de vue acoustico-phonétique, des effets de la maladie neurologique sur la production de la parole, enrichissant pour le diagnostic des cliniciens, mais également à des fins d'enseignement des troubles de la voix destinés aux orthophonistes et autres spécialistes de la voix.

Dans une perspective future, il est évident que les résultats obtenus doivent être validés sur une plus large base de données pathologique afin d'augmenter la fiabilité des résultats. Cette expérience nous permet cependant, de dire que la méthode, utilisant comme caractéristiques acoustiques le Jitter, le Shimmer et l'énergie, nous a donné une performance appréciable avec notamment une précision de reconnaissance de 92,50% des VN et 95,00% des VP. Le Taux global de reconnaissance TGR de l'ensemble des fichiers sons se situe autour de 93,75%.

En conclusion, nous dirons que le système de classification automatique utilisé donne des résultats assez intéressants. Il reste que d'autres paramètres peuvent contribuer sensiblement à l'amélioration du TGR. Nous pensons notamment au HNR (Harmonics to Noise Ratio) qui représente un paramètre assez pertinent dans l'analyse acoustique des paroles et voix pathologiques. Ce paramètre doit être mis au point au moyen du langage de programmation Matlab® et intégré à notre processus de classification automatique.

Bibliographie

1. Ananthakrishna, T., Shama, K., & Niranjana, U. C. (2004). k-means nearest neighbour classifier for voice pathology. *Proceedings of the IEEE India Conference (INDICON)* (pp. 352–354). Indian Institute of Technology, Kharagpur, India.
2. Ben Messaoud, [M. A.](#), Bouzid, [A.](#) (2009). A New Method for Pitch Tracking and Voicing Decision Based on Spectral Multi-Scale Analysis. *Signal Processing: An International Journal (SPIJ)*, 3(5):83-171.
3. Childers, D. G., Skinner, D. P., & Kemerait, R. C. (1977). The Cepstrum: A Guide to Processing, *Proceedings of the IEEE*, 65, pp.1428-1443.
4. Dev, A., Bansal, P. (2010). Robust Features for Noisy Speech Recognition using MFCC Computation from Magnitude Spectrum of Higher Order Autocorrelation Coefficients. *International Journal of Computer Applications*, 10, pp.36-38.
5. Dreyfus, G. (2004). *Réseaux de neurones- Méthodologie et Application*. Ed. Eyrolles, Paris, France.
6. Ferrat K., Guerti M. (2011). Apprentissage et Reconnaissance Automatique de la Parole par Réseaux de Neurones Artificiels. *Revue Sciences de l'homme*, Laboratoire SLANCOM, Université d'Alger2, N°4, 2011, pp.57-71.
7. Foresee, F.D., Hagan, M.T. (1997). Gauss-Newton Approximation to Bayesian Regularization. *Proceedings of the 1997 International Joint Conference on Neural Networks (IJCNN'97)* (pp. 1930-1935). Houston, Texas, USA.
8. Fun, M.H., Hagan, M.T. (1996). Levenberg-Marquardt Training for Modular Networks. *Neural Networks, IEEE International Conference on*, 1, pp.468 – 473.
9. Godino-Llorente, J. I., & Gómez-Vilda, P. (2004). Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Transactions on Biomedical Engineering*, 51, pp.380–384.
10. Holmes, R. J., Oates, J. M., Phyland, D. J., & Hughes, A. J. (200). Voice characteristics in the progression of Parkinson's disease. *International Journal of Language & Communication Disorders*, 35, 407 – 418.
11. Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry*, 79, pp.368–376.
12. Kreiman, J., & Gerratt, B. R. (2005). Perception of aperiodicity in pathological voice. *J. Acoust. Soc. Amer.*, 117, pp.2201-2211.

13. Klingholz, F. & Martin, F. (1985). Quantitative spectral evaluation of shimmer and jitter. *Journal of Speech and Hearing Research*, 28, pp.169-174.
14. Lachmann, T., Schumacher, B., Joebges, M., Hummelsheim, H., & van Leeuwen, C. (2008). Procedural learning eliminates specific slowing down of response selection in patients with idiopathic Parkinson syndrome. *Journal of Clinical and Experimental Neuropsychology*, 30, pp.319-326.
15. Locco, J. (2005). *La production des occlusives dans la maladie de Parkinson*. Thèse de Doctorat de l'Université de Provence. Aix-Marseille I, France.
16. Logeman, J.A., Fisher, H.B., & Boshes, B. (1978). Frequency and co-occurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *Journal of Speech and Hearing Disorders*, 43, pp.47-57.
17. Mitchell, R. L. C., & Bouças S. B. (2009). Decoding emotional prosody in Parkinson's disease and its potential neuropsychological basis. *Journal of Clinical and Experimental Neuropsychology*, 31, pp.553-564.
18. Povinelli, R. J., Johnson, M.T., Lindgren, A. C., & Ye, J. (2004). Time Series Classification Using Gaussian Mixture Models of Reconstructed Phase Spaces. *Knowledge and Data Engineering, IEEE Transactions on*, 16, pp.779-783.
19. [Ramig, L.O.](#), [Fox, C.](#), [Sapir, S.](#) (2008). Speech treatment for Parkinson's disease. *Expert Rev Neurother*, 8, pp.297-309.
20. Rigaldie, K., Nespoulous, J.L., Vigouroux, N. (2004). Dysprosody in Parkinson's disease: An acoustic study based on tonal phonology and the INTSINT system. *International Conference on Speech Prosody* (pp. 617-620). Nara, Japan.
21. Ritchings, R.T., Mcgillion, M.A., & Moore, C.J. (2002). Pathological voice quality assessment using artificial neural network, *Medical Engineering & Physics*, 24, pp.561-564.
22. Sapir, S., Ramig, L., & Fox, C. (2006). The Lee Silverman Voice Treatment [LSVT®] for Voice, Speech, and Other Orofacial Disorders in People with Parkinson's disease. *Future Neurology*, 1, pp.563-570.
23. Shimamura, T., & Kobayashi, H. (2001). Weighted Autocorrelation for Pitch Extraction of Noisy Speech. *IEEE Transactions on Speech and Audio Processing*, 9, pp.727-730.
24. Shimamura, T., & Takagi H. (2004). "Noise-Robust Fundamental Frequency Extraction Method Based on Exponentiated Band-Limited Amplitude Spectrum". *47th IEEE International Midwest Symposium on Circuits and Systems* (pp. II141-II144). Hiroshima, Japan.
25. Skodda, S., & Schlegel, U. (2008). Speech rate and rhythm in Parkinson's disease. *Journal of Movement Disorders*, 23, pp.985-992.
26. Sun, X. (2002). Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio. *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, (pp. I333-I336). Orlando, Florida, USA.
27. Terez, D. (2002). Fundamental frequency estimation using signal embedding in state space. *Journal of the Acoustical Society of America*, 112, pp.2279-2279.
28. Tiwari, V., (2010). MFCC and its applications in speaker recognition. *International Journal on Emerging Technologies*, 1, pp.19-22.
29. Umaphathy, K., Krishnan, S., Parsa, V., & Jamieson, D.G. (2005). Discrimination of pathological voices using a time-frequency approach. *IEEE Transactions on Biomedical Engineering*, 52, pp.421-430.
30. Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. (1989). Phoneme recognition using time-delay networks. *IEEE Trans. Acoustics, Speech and Signal Processing*, 37, 328-339.

31. Wang, J., & Jo, C. (2006). Performance of Gaussian Mixture Models as a classifier for pathological voice. *Proceedings of the 11th Australasian International Conference on Speech Science and Technology* (pp.165-169). University of Auckland, New Zealand.
32. Werbos, P.J. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, 78, 1550–1560.
33. Zellal, N. (1984). Introduction à la phonétique orthophonique arabe, éditions O.P.U, Alger, 1984.