

DOSSIER DE CANDIDATURE AU PRIX DE THESE IFRATH 2015

M. Mohammed ZBAKH

Docteur en Informatique

Directeur de thèse : Jaime LOPEZ KRAHE

Rapporteurs :

Christian BERGER-VACHON, Professeur, Université Lyon 1

Alain PRUSKI, Professeur, Université de Lorraine

Examineurs :

Dominique ARCHAMBAULT, Professeur, Université Paris 8

Christian CUXAC, Professeur, Université Paris 8

Cédric MOREAU, Maître de Conférences, INSHEA, Suresnes

Catherine PELACHAUD, Directrice de recherche, CNRS

Contenu du document

| | |
|---------------------------|---------|
| Curriculum Vitae détaillé |3 |
| Résumé de la thèse |5 |
| Liste des publications |8 |
| Avis du directeur |9 |
| Rapports de thèse |10 |
| Rapport de soutenance |15 |
| Article 1 |16 |
| Article 2 |28 |
| Article 3 |37 |

Curriculum Vitae Détaillé

Prénom et Nom : Mohammed ZBAKH
Date et lieu de naissance : 13-08-1981 à Tétouan – Maroc
Adresse personnelle : 31 Allée du Maréchal Mortier, 93270 Sevran
Adresse professionnelle : 2 rue de la liberté, 93526 Saint-Denis Cedex
Téléphone bureau : 01 49 40 73 47
Téléphone portable : 06 85 26 42 95
E-mail : mohammed.zbakh02@univ-paris8.fr

Formation et Diplômes

2014 : Doctorat, Spécialité : Informatique (Cognition)
Thème : Apports du numérique dans les outils de communication des personnes handicapées - Développement d'un dictionnaire inversé : Langue des Signes Française → Français
Laboratoire : L'équipe THIM (Technologie, Handicap Interfaces et Multimodalités), du laboratoire CHART EA4004 (Cognition Humaine et ARTificielle)
Directeur de thèse : Jaime LOPEZ KRAHE
2009 : Master 2, Technologies et Handicap à l'Université Paris8 de Saint-Denis.
2008 : Master 1, Technologies et Handicap à l'Université Paris8 de Saint-Denis.
2007 : Maîtrise, IUP NTIC Intégration des Systèmes Informatiques à l'Université Pascal PAOLI de Corse.
2007 : Ingénieur, en Systèmes et Réseaux Informatiques à l'école eHECT, Tanger.

Expériences professionnelles

Sep. 2012 – Aout 2015 : Ingénieur d'étude et de développement au laboratoire THIM

- Portail web pour un dictionnaire inversé de la LSF.
- Application android d'aide à la mobilité dans les grandes gares de la SNCF.
- Plateforme web pour l'exploration des sceaux historiques en utilisant un bras haptique

Sep. 2010 – Aout 2011 : Attaché Temporaire d'Enseignement et de Recherche (ATER)

- Enseignement : - Algorithmique, Langage de programmation C++, Conception et méthode d'analyse MERISE, Programmation Web : PHP4, PHP5 & MySql, et Traitement de signal et reconnaissance de formes.
- Encadrement : Projets techniques

Sep. 2011 – Aout 2012 : Attaché Temporaire d'Enseignement et de Recherche (ATER)

- Enseignement : - Algorithmique, Langage de programmation Python, Conception et méthode d'analyse MERISE, Programmation Web : PHP4, PHP5 & MySql et Traitement de signal et reconnaissance de formes.
- Encadrement : Projets techniques

Nov. 2009 – Juil. 2010 : Ingénieur d'étude et de développement au laboratoire THIM

- Adaptation et intégration d'un module du Langage Parlé Complété (LPC) dans l'agent conversationnel Greta

Compétences techniques

Systèmes d'exploitation : Windows, Unix, Linux

Programmation web : PHP5, XML/XSL, CSS, HTML, Javascript, jQuery, Apache, AJAX, Drupal, Wordpress

Bases de données / Outils décisionnels: MySQL, SQL Server, SQLite, Oracle, MS Access, PostgreSQL, Crystal Reports

Langages : Visual Basic, C/C++, Java, Python, MatLab

Réseaux : Ethernet, TCP/IP, Routeur CISCO, VLAN, VPN, DHCP, FTP

Modélisation objet : UML, Merise

Graphisme : PhotoShop, DreamWeaver, Ulead Media Studio

Outils : WAMP, Visual Studio, C++ Builder, Dev C++, JCreator, NetBeans, MS Project, OpenOffice, WinSCP, PowerAMC

Langues

Français Courant

Anglais Niveau technique

Espagnol Niveau A3 (Intermédiaire)

LSF Niveau 2

Allemand Notions de base

Résumé de la thèse

Thèse de Doctorat

Titre : Apports du numérique dans les outils de communication des personnes handicapées - Développement d'un dictionnaire inversé : Langue des Signes Française → Français

Mots clés : Classification – Apprentissage – Langue des signes française – Interface Homme Machine

Date de début : Septembre 2009

Date de soutenance : 17 Décembre 2014

Rapporteurs :
Christian BERGER-VACHON, Professeur, Université Lyon 1
Alain PRUSKI, Professeur, Université de Lorraine

Examineurs :
Dominique ARCHAMBAULT, Professeur, Université Paris 8
Christian CUXAC, Professeur, Université Paris 8
Cédric MOREAU, Maître de Conférences, INSHEA, Suresnes
Catherine PELACHAUD, Directrice de recherche, CNRS

Contexte de la thèse

Travaux préalables

J'ai commencé ma thèse par une étude sur les outils techniques de communication pour les personnes handicapées. Le premier travail par lequel j'ai débuté ma recherche est une application d'aide à la communication par le biais des pictogrammes pour des personnes dysphasique. C'était un projet de master que j'ai complété par la suite en ajoutant des nouvelles fonctions et en améliorant la traduction des phrases pictographiques en phrases textuelles accompagnée d'une synthèse vocale.

Deuxième travail consistait à travailler sur un agent conversationnel capable de s'exprimer verbalement et non verbalement. L'objectif est de développer un module du langage parlé complété est de l'intégrer dans un agent conversationnel nommé Greta. Il s'agit d'une interprétation d'une entrée textuelle en codage LPC accompagné de mouvements de lèvres.

Dictionnaire inversé

L'équipe THIM (Technologie, Handicap Interfaces et Multimodalités), du laboratoire CHART EA4004 (Cognition Humaine et ARTificielle), a développé ces derniers temps une base de données lexicale en LSF (Langue des Signes Française) sur le Web. Ce projet, réalisé en collaboration avec l'Académie de la Langue des Signes (ALSF), soutenu par l'aide de la Région Ile de France dans le cadre des projets PICRI (Partenariat Institutions Citoyens pour la Recherche et pour l'Innovation), nous a permis le développement d'un portail WEB en libre accès qui comporte aujourd'hui presque 3000 vidéos de terminologie en LSF.

Problématique

L'accès actuel est exclusivement dans le sens Français -> LSF. Par contre, le processus inverse n'est pas possible : trouver le mot à partir du signe pose des problèmes théoriques plus complexes et il est impossible de trouver la signification d'un signe à partir du signe lui-même ou d'une approximation de celui-ci. Etant donné que la langue des signes est une langue gestuelle mettant en relation plusieurs paramètres, les chercheurs et les auteurs de cette langue ne se sont pas toujours mis d'accord sur un protocole standard pour définir ses paramètres. A titre d'exemple, il existe toujours un désaccord sur le nombre défini de configurations de la main.

L'objectif de mes travaux de thèse est de développer un outil simple et accessible capable de rendre la recherche des signes par ses paramètres une opération possible pour toute personne qui s'intéresse à cette langue.

Réalisation

Un signe est défini par cinq paramètres : une configuration de la main, un emplacement dans l'espace, un mouvement, une orientation de la main et des expressions faciales. En revanche, demander à un utilisateur de saisir cinq paramètres pour trouver un signe provoque une charge mentale très importante. Pour cela, j'ai limité, par des contraintes ergonomiques, le nombre des paramètres saisis lors de la recherche. Puis, j'ai rattrapé le manque de ces informations par l'ajout d'un filtreur de nombre de mains et l'utilisation d'algorithmes d'apprentissage dynamique.

Après avoir défini ergonomiquement les paramètres utilisés dans ce travail et le nombre des éléments de chaque paramètre, j'ai commencé mon travail par la création d'une base de données des signes tout en affectant un codage spécifique pour chaque signe par les paramètres définis. Ensuite j'ai initialisé ces paramètres par des matrices de dissimilarité pré-remplies par des pratiquants de la LSF. Mon système a été testé dans un premier temps pour calculer le poids de chaque paramètre dans l'espace de représentation des signes afin de définir le paramètre le plus discriminant pour différencier les signes. Cette opération a été complétée par un programme extérieur que j'ai développé pour jouer le rôle de testeur. Il s'agit d'une réexécution en boucle des codages des utilisateurs avec un changement périodique des valeurs des poids pour chaque paramètre. A cette étape, mon système se base sur la fonction de calcul suivante :

$$DIS(S_{ref}, S_{user}) = \sum_{i=1}^2 \alpha (dis(C_{ref(i)}, C_{user(i)})) + \beta (dis(P_{ref(i)}, P_{user(i)})) + \gamma (dis(M_{ref(i)}, M_{user(i)})) + \delta (dis(N_{ref(i)}, N_{user(i)}))$$

Avec :

- $DIS(S_{ref}, S_{user})$: La dissemblance globale entre le signe et le codage de l'utilisateur.
- $dis(C_{ref(i)}, C_{user(i)})$: La dissemblance entre les configurations.
- $dis(P_{ref(i)}, P_{user(i)})$: La dissemblance entre les emplacements.
- $dis(M_{ref(i)}, M_{user(i)})$: La dissemblance entre les mouvements.
- $dis(N_{ref(i)}, N_{user(i)})$: La dissemblance entre les nombre de la main.
- $\alpha, \beta, \gamma, \delta$: Ce sont des coefficients d'apprentissage avec $\alpha + \beta + \gamma + \delta = 1$

Nous avons constaté que cette méthode ne donne pas des bons résultats sauf si l'utilisateur est un habitué de la langue des signes. Les autres utilisateurs novices en LSF ont rencontré des difficultés

lors de la recherche d'un signe à cause de l'ambiguïté causé par certains paramètres. Pour répondre à cette question, nous avons fait appel à des algorithmes de classification classiques comme les réseaux de neurones et les arbres de décision. Ces derniers n'ont pas réalisé un taux de réussite acceptable pour les intégrer dans le système. Pour remédier à cela, j'ai créé un tableau d'apprentissage statistique basé sur les erreurs des utilisateurs. Il s'agit de suivre les scénarios des codages utilisateurs et analyser le comportement de chaque paramètre durant la phase de codage. Le tableau d'apprentissage présente les correspondances de tous les codages de la base de données avec les signes possibles pour chaque codage selon les expérimentations des utilisateurs. Enfin, j'ai calculé la probabilité d'un signe pour un codage donné pour afficher le résultat sous la forme d'un nuage de mots capable de trouver le signe exacte ou proche suivant le choix de l'utilisateur.

La dernière partie de mon travail consiste à la création d'une méthode automatique et dynamique pour enrichir le dictionnaire de la LSF. J'ai mis en place un système d'auto-enrichissement basé sur trois moteurs : un moteur d'extraction capable d'analyser et d'identifier les profils des utilisateurs pour attribuer un degré de confiance pour chacun, un moteur de validation qui permet de corriger et de valider les nouveaux codages des utilisateurs et un moteur d'enrichissement pour insérer les nouveaux signes dans la base de données.

Perspectives :

L'augmentation de l'espace de représentation des signes produira plus d'ambiguïté sur les signes quand la recherche se base sur trois paramètres. Actuellement je travaille sur l'amélioration des critères de recherche par l'intégration du principe de l'iconicité comme un paramètre de détection des signes.

Liste des publications

Revues

M. **Zbakh**, Z. Haddad, J. Lopez Krahe, A cognitive approach for signs classification Development of an online French Sign Language dictionary, Pattern Recognition Letters, Vol 67, pp. 28-38, December 2015.

M. **Zbakh** - H. Daassi-Gnaba - J. Lopez Krahe, "*Talking head generating French cued speech for deaf and hard of hearing people*", accepté dans la revue AMSE (association for the Advancement of Modelling and Simulation techniques in Enterprises), Vol. 71 n°3, pp. 117,127, 2011.

H. Daassi-Gnaba. - M. **Zbakh** - J. Lopez Krahe, "*Combinaison de reconnaissance de la parole, reconnaissance des émotions et tête parlante codeuse en LPC pour les personnes sourdes et malentendantes*", STH (Sciences et Technologies pour le Handicap), Vol 3, pp. 239-253, Hermès Lavoisier, 2010.

Conférences

M. **Zbakh**, Z. Haddad, J. Lopez Krahe "*Toward a reversed dictionary of French Sign Language (FSL) on the Web*", ICCHP 2014, pp. 423,430, Paris, 9-11 Juillet 2014

M. **Zbakh**, J. Lopez Krahe, "*Dictionnaire Inversé : Langue des Signes Française –Français* ", HANDICP 2012, pp. 46-50, Paris, Juin 2012.

M. **Zbakh**, "*Dictionnaire Inversé : Langue des Signes Française –Français* ", JCJC 2011, pp. 93-102, Paris, Juin 2011.

M. **Zbakh**, J. Lopez Krahe, "*Reversed dictionary : French Sign Language - French*", DRT4ALL, pp. 20,27, Madrid,27-29 Juin 2011.

M. **Zbakh** - H. Daassi-Gnaba - J. Lopez Krahe, "*Tête parlante codeuse en LPC pour les sourds et les malentendants* ", HANDICP 2010, pp. 16-21, Paris, Juin 2010.

M.**Zbakh**, I.Lopez Fontana, K.Anache, A.Mortera, J.Lopez Krahe, "*Pictokids : un logiciel de communication pictographique avec sortie textuelle ou vocale* ", HANDICP 2010, pp. 155-162, Paris, Juin 2010.

Prix

Pictokids : 1er Prix de communication aux challenge Handicap et Technologies à Metz, 2009

Dictionnaire inversé : 1er Prix de communication aux challenge Handicap et Technologies à Paris, 2013



UFR TES
Domaine : Sciences, Technologie, Santé
Master Ingénierie et Cognition
Spécialité : Technologie et Handicap
THIM : Technologie, Handicaps, Interfaces, Multimodalités



Lettre de recommandation

M Mohammed ZBAKH a effectué une thèse sous ma direction dont le titre est : « *Apports du numérique dans les outils de communication des personnes handicapées. Développement d'un dictionnaire inversé : Langue des Signes Française (LSF) → Français* ». Cette thèse a été soutenue le 17 décembre 2014 avec la mention très honorable.

Les dictionnaires Français ->LSF sont apparus sur internet depuis une dizaine d'années. Le vocabulaire LSF est contenu dans une base de données vidéo. A partir d'un thème ou d'un terme de vocabulaire on peut visualiser le signe LSF, ce qui peut constituer une aide en ligne pour l'apprentissage du vocabulaire, par exemple.

Le processus contraire, c'est-à-dire un dictionnaire LSF -> français, est plus complexe. Le premier problème à résoudre est de trouver des paramètres qui permettent une description du signe, d'une manière rapide et ergonomique tout en ayant les éléments nécessaires pour définir le signe que l'on cherche. Ceci n'est pas simple et très peu de travaux ont été développés sur ce type de transcription. Ensuite il faut gérer les approximations, détecter les erreurs de description et mettre en place un système d'apprentissage statistique.

Le sujet de la thèse, très original, porte principalement sur la conception d'un dictionnaire inversé LSF-> français. M Zbakh a abordé avec constance et énergie ce travail difficile, cherchant plusieurs voies de codage et classification, établissant les paramètres les plus adaptés et discriminants, développant les interfaces utilisateur, lançant de nombreuses expérimentations, ce qui a autorisé la création d'une importante base de données d'apprentissage (plusieurs milliers d'éléments). Ceci a permis de prendre en compte les erreurs les plus fréquentes des usagers et d'établir ainsi un système de reconnaissance très performant et adaptatif. L'application est opérationnel et peut être consulté librement sur internet : <http://www2.univ-paris8.fr/ingenierie-cognition/master-handi/di/>

Le sujet est très bien adapté au prix de thèse convoqué par l'IFRATH et pour le volume de travail et les résultats obtenus je soutiens sans réserve sa candidature.



THIM
EA 4004 CHART
Technologie Handicap
Interfaces et Multimodalités
Université Paris 8

Rapports de thèse



Rapport sur le mémoire de thèse de Doctorat de
Mohammed ZBAKH
intitulé

**Apports du numérique dans les outils de communication des personnes handicapées.
Développement d'un dictionnaire inversé : Langue des Signes Française vers le Français".**

Le mémoire est composé de 8 chapitres avec une introduction, une conclusion et des annexes. La thèse est divisée en deux parties.

L'introduction présente le sujet de la thèse qui concerne la communication humaine et plus particulièrement les problèmes relatifs aux personnes sourdes et à la langue des signes.

La première partie traite de "La communication et la surdité : obstacles et solutions", elle est composée de trois chapitres et repose sur deux réalisations qui constituent une partie exploratoire de la thèse.

La deuxième partie "Le dictionnaire inversé: un outil d'accès à la LSF" est composée de cinq chapitres et constitue la partie la plus développée de la thèse

Première partie

Le premier chapitre traite de la communication humaine, du langage et des obstacles de la communication, en explorant plus spécifiquement les troubles cognitifs et auditifs. Il évoque les solutions pictographiques, le langage parlé complété (LPC), la langue des signes et les implants cochléaires. Il s'agit d'un premier chapitre classique d'introduction des sujets qui seront abordés dans les deux chapitres suivants.

Le chapitre deux présente un travail qui avait été initialisé dans le cadre d'un projet de groupe du master Technologie et Handicap et qui a été poursuivi par l'auteur en début de sa thèse pour aboutir à un produit fini. Il s'agit de pallier les problèmes de communication suite à des pathologies cognitives (dysphasie) ou temporelles suite à un accident, par exemple. A partir d'une communication par pictogrammes dont le choix est hiérarchisé (thème, sous-thème, liste) ou alphabétique, le système propose une sortie textuelle ou vocale après un traitement grammaticale basique mais efficace. La question ergonomique semble avoir été bien traitée et une évaluation a été effectuée avec des résultats très encourageants. L'application est mise à disposition sur le site internet du master Technologie et Handicap mais il serait intéressant qu'elle soit aussi adaptée aux smart phones pour une plus grande diffusion.

Le chapitre trois traite du développement d'une tête parlant codeuse en LPC comme élément pouvant aider les personnes sourdes capables réaliser une lecture labiale et pratiquant cette méthode de codage. Ce système, assez récent, a été développé pour permettre de lever les ambiguïtés des phonèmes en lecture labiale qui ont la même perception visuelle. M Zbakh a travaillé sur l'Agent

Conversationnel Animé (ACA) "Greta", en collaboration avec le département TSI de Télécom Paris Tech, pour lui permettre de communiquer en LPC, en synchronisant les mouvements des lèvres de la tête parlante et la gestualité du LPC. Dans les tests effectués, la programmation du système LPC semble fonctionner correctement, cependant des problèmes de synchronisation en temps réel entre la synthèse de parole et le mouvement des lèvres de l'ACA ne semblent pas avoir été résolus par l'auteur probablement pour une question de vitesse de calcul. Ce sujet est intéressant et a donné lieu à d'autres travaux qui sont cités dans la bibliographie, en particulier les études de l'expression faciale à partir de l'analyse émotionnelle de la voix et de la reconnaissance de la parole.

Deuxième partie

La deuxième partie est dédiée à l'étude d'un dictionnaire inversé de LSF vers le français. Le chapitre quatre introduit la problématique du dictionnaire inversé. Des dictionnaires lexicaux du français vers la LSF existent sur Internet depuis une dizaine d'années et quelques liens sont présentés ici. Il est possible d'obtenir une séquence vidéo en LSF correspondant à un mot en français. Le problème inverse est plus compliqué puisque le nombre de paramètres qui permet de décrire un signe est évoqué montrant que le travail de recherche peut être extrêmement lourd pour rendre le dictionnaire utilisable. Un compromis semble nécessaire pour aller du signe vers le mot.

Le chapitre cinq est intitulé "Analyse de l'espace de représentation et modélisation des paramètres de la LSF". Le projet se concentre sur un vocabulaire réduit de 500 mots sélectionnés en fonction des occurrences des demandes du site de Paris 8 qui est utilisé comme dictionnaire de référence. Ce vocabulaire pourra être augmenté par la suite dans le cadre d'un travail collaboratif (crowdsourcing). Un premier espace de représentation sur deux paramètres (configuration de la main et position dans l'espace) qui s'inspire d'un travail antérieur est utilisé pour effectuer les premiers tests sur des mesures de dissimilarité dans l'espace de représentation. En fonction des résultats obtenus, des paramètres sont modifiés (position dans l'espace) et d'autres sont ajoutés (mouvements de ou des main(s), présence d'une ou deux mains) tout en restant dans un contexte d'emploi simple. L'espace de représentation réduit produit quelques homosignes, qui auront le même code, ce qui ne se traduit pas obligatoirement par des signes identiques dans la réalité de la LSF. Une interface de recherche d'un signe est développée et la présentation des résultats peut se faire soit sous la forme d'une liste ordonnée en fonction de la similarité avec les codes de référence, soit sous le format d'un "mur" (plus "convivial") dont la taille des mots est en rapport avec une évaluation de la similarité.

Le chapitre six est dédié aux critères de classification et aux expérimentations. L'auteur a constitué une base de données de 7230 observations expérimentales (mots codées par 36 participants de divers niveaux en LSF) dont 6775 validées pour l'étude. Une analyse des résultats à partir des paramètres de manière indépendante permet de se faire une idée de l'influence globale dans le signe. L'auteur teste quelques méthodes d'apprentissage et de reconnaissance (réseaux de neurones, arbre de décision) mais s'oriente vers un système de coefficients d'apprentissage sur un échantillon aléatoire de 600 observations. Les résultats s'obtiennent dans la reconnaissance des autres observations. Dans une troisième expérimentation est étudiée l'influence du niveau de pratique de la LSF dans les résultats obtenus en vue d'adapter la plateforme au niveau de chaque utilisateur. L'auteur propose l'utilisation des erreurs des participants pour mettre en place un apprentissage adaptatif et pour améliorer les performances.

La base des données obtenue semble très intéressante et il serait bon qu'elle puisse être publique afin que d'autres chercheurs puissent travailler sur les mêmes données et comparer les résultats.

Le chapitre sept intitulé "Apprentissage des signes et traitement des profils des utilisateurs" exploite les résultats évoqués dans le chapitre antérieur. Après avoir analysé les erreurs des participants, l'auteur a pu constater que ces erreurs sont souvent systématiques, en particulier dans le cas des apprentis de la LSF. Par exemple, dans le cas d'un signe à double configuration, il est fréquent que seul l'enregistrement de la configuration initiale ou finale du signe soit réalisé, ce qui provoque une erreur de reconnaissance par comparaison à la configuration de référence. Le taux de reconnaissance après l'apprentissage de ces erreurs augmente de manière considérable : 61% contre 39% pour la première réponse et 96% contre 64% sur le cumul de 5 premières réponses (tableaux 6.6 et 7.3).

Ces résultats, originaux, mènent l'auteur à considérer de manière pertinente qu'ils peuvent être utilisés dans le cadre d'une application pédagogique personnalisée. Il serait intéressant d'approfondir cette voie qui semble très prometteuse.

Le chapitre huit intitulé " Enrichissement automatique du dictionnaire inversé" propose une méthode d'auto-enrichissement basée sur une approche de crowdsourcing (collaboration des utilisateurs). Un moteur d'extraction associé à un moteur de validation en fonction du degré de fiabilité des utilisateurs. Les critères de validation sont bien décrits, la méthode semble originale et devrait permettre un enrichissement de la base de référence à faible coût.

Un chapitre de conclusions et perspectives qui résume l'état actuel de ces travaux de recherche et les voies et perspectives clos le manuscrit.

De nombreuses références sont citées dans la bibliographie dont six publications du candidat dans des conférences à comité de lecture ainsi que dans une revue internationale. Il serait judicieux que l'auteur publie les derniers résultats qui ne semblent pas être cités dans les références, en particulier ceux qui correspondent aux chapitres 7 et 8.

Le travail dont il est fait mention dans ce manuscrit est conséquent au niveau du volume attendu d'une thèse. La première partie aurait pu être éludée. Elle correspond à la poursuite d'un travail dans le cadre d'un master en équipe. La seconde partie relève d'une cohérence globale et constitue un volume de travail tout à fait suffisant et original. Le manuscrit souffre encore de quelques fautes de frappe qu'il faudra corriger avant la publication du document final. Pour toutes ces raisons j'émet un avis favorable à la soutenance de la thèse.

Fait à Metz, le 13 décembre 2014



Alain Pruski
Prof. à l'Université de Lorraine

**Rapport sur le travail présenté par Monsieur Mohammed Zbakh pour l'obtention d'une
thèse de doctorat en informatique de l'Université Paris 8-Vincennes-St-Denis
(Ecole doctorale Cognition, Langage, Interaction)**

M. Zbakh a abordé un problème important qui est celui du décodage de la langue des signes. On sait que le passage par des interprètes de la langue des signes est très coûteux et il convient de simplifier cette approche lorsque ceci est possible. Un dictionnaire LSF → Français mérite donc d'être considéré.

Le mémoire présenté par M. Zbakh se compose de huit chapitres qui peuvent être résumés ci-dessous.

Chapitre 1 : Il s'agit de rappels sur la communication, sur la LSF (Langue des Signes Française). A chaque codage gestuel correspond un signe, mais l'affaire est beaucoup moins simple, car les gestes sont parasités par la variation liée à toute activité humaine et l'interprétation d'un codage n'est pas univoque.

La main joue un rôle fondamental dans ce codage et l'information est portée par la configuration, la place, le mouvement, le nombre de mains concerné et aussi par l'orientation même si ce caractère a été peu développé dans ce travail.

Chapitre 2 : le chapitre 2 est consacré à Pictokids, système de communication non verbale développée par l'auteur, qui tient une place de précurseur. Il présente un mode de communication imagée où le support est l'image. Une organisation logique permet de renforcer l'efficacité des échanges, à travers des notions grammaticales et thématiques. Ce chapitre montre la structure logique d'un mode de communication, mais semble toutefois un peu marginal pour la suite du travail qui est consacré à la LSF.

Chapitre 3 : Il introduit la tête codeuse Greta, un avatar que l'auteur a développé pour l'adapter à communiquer avec le LPC (Langage Parlé Complété) ; la main vient compléter pour faciliter la compréhension des signes labiaux. L'avatar aide à caractériser les voyelles et les consonnes. Ce codage s'affranchit du mouvement des lèvres. A priori, cette représentation s'approche des notions qui seront développées par la suite dans le cadre du projet du dictionnaire inversé LSF → Français permettant de passer d'un codage au signe qui lui est associé pour trouver l'équivalent en français.

Chapitre 4 : Le passage d'un codage à un signe demande à être précisé. Le codage des signes par la main s'appuie sur quatre paramètres essentiellement, la configuration de la main, la position de la main par rapport au corps, le mouvement de la main, et l'emploi d'une ou deux mains pour coder. On a donc affaire à un espace à 4 dimensions pour caractériser un signe. Cette partie est fondamentale pour comprendre la suite du travail.

Chapitre 5 : Ce chapitre présente une description plus précise des paramètres. Cinquante-neuf éléments sont proposés pour la configuration de la main. Certains sont proches les uns des autres et d'autres sont très différents. Une étude des confusions (dissemblances) permet de préciser cet aspect, et ensuite de regrouper les éléments similaires en classes.

Un travail similaire est proposé pour la position de la main dans l'espace et ceci est ensuite simplifié pour limiter « l'explosion combinatoire » (27 possibilités sont ramenées à 15). Enfin la représentation des signes fait appel au mouvement et onze nouvelles possibilités viennent s'ajouter au système.

Le codage concerne ensuite 2234 signes vidéo et 500 (les plus courants) ont été analysés.

Chapitre 6 : Une fois qu'un codage est caractérisé, il faut trouver les signes correspondant et donc interroger la base de données pour voir le signe « le plus proche ». Au même titre qu'il y a en lecture labiale « des sosies labiaux », on trouve ici des « homosignes » qu'il faudra distinguer « in fine ».

Une autre composante est aussi à prendre en compte « la qualité du signage » et les habitudes propres à chaque signeur. Ces variabilités, inhérentes à l'être humain, seront à prendre en compte dans la décision.

Des méthodes de décisions classiques ont été considérées :

- les réseaux de neurones
- l'arbre de décision.

Ces méthodes s'avèrent peu adaptées au problème posé et finalement l'étude à l'aide des matrices de dissemblance semble conduire aux meilleurs résultats. Cette dernière approche mériterait d'être mieux précisée, notamment le tableau 6.6 qui s'appuie sur la représentation de l'espace à 4 dimensions, dimensions qui peuvent être pondérées (métriques classiques).

De même, la significativité des résultats indiqués p 121 entre différents types de participants est-elle significative ? Ceci mérite d'être indiqué pour évaluer la robustesse de la méthode de classement.

Chapitre 7 : Comme on l'a déjà indiqué, le profil du signeur a son importance et l'augmentation du nombre d'utilisateurs accroît le nombre de codages proposés. A cet effet, le tableau 7.3 mériterait d'être mieux détaillé, notamment en précisant l'expérience qui a conduit à ces résultats. Il est clair que la variabilité du codage est un point important à prendre en compte, surtout si on veut que le système proposé ait des retombées pédagogiques.

Chapitre 8 : Enfin, il est à peu près évident que la méthode proposée ne peut pas être figée et elle doit pouvoir s'enrichir pour tenir compte du profil des utilisateurs et même personnaliser les codeurs. La possibilité qui est offerte à cet effet est intéressante et elle rendra service, notamment aux débutants.

En conclusion, Le travail proposé est intéressant et important, compte tenu du nombre de signes proposé et des possibilités de codage qu'il fallait analyser. Le système demande à être étoffé dans le futur et il doit prendre sa place dans les moyens de communication avec la Communauté des Sourds. La bibliographie est correcte, (61 références) mais elle pourrait encore être étoffée.

En ce qui concerne les publications du candidat, on note des présentations en congrès (soumis à des comités de lecture exigeants) et un article en revue. Ceci devra aussi être étoffé dans le futur.

Dans ces conditions, je pense que le travail de Mohammed Zbach peut être présenté à l'université Paris 8-Vincennes-St Denis pour l'obtention du grade de docteur en informatique au sein de l'école doctorale « Cognition, Langage, Interaction ».

Le 4 octobre 2014

Christian Berger-Vachon
Professeur Émérite 61^e section
Université Lyon 1

Rapport de soutenance de thèse

Rapport de soutenance de thèse

Auteur : Mohammed Zbakh

Domaine : Sciences, Technologie, Santé

Discipline : INFORMATIQUE

ED 224 : Cognition, langages, interaction (CLI)

Titre : « *Apports du numérique dans les outils de communication des personnes handicapées. Développement d'un dictionnaire inversé : Langue des Signes Françaises → Français* »

Soutenue à l'Université Paris 8 le 17 décembre 2014 devant un jury composé de :

Président : Pr Dominique Archambault - Université Paris 8-Vincennes-Saint-Denis

Rapporteurs : Pr Christian Berger-Vachon, Université Claude Bernard, Lyon I
Pr Alain Pruski, Université de Lorraine

Examineurs : Pr Christian Cuxac, Université Paris 8, Dr Cédric Moreau, Maître de Conférences INSHEA, Catherine Pélachaud, Directrice de recherches CNRS, Telecom ParisTech.

Directeur : Pr. Jaime Lopez Krahe, Université Paris 8-Vincennes-Saint-Denis

Monsieur Mohammed Zbakh a soutenu sa thèse devant un jury à forte composante multidisciplinaire (Informatique, Linguistique, Traitement du signal). Il a fait une présentation claire, pédagogique, et bien illustrée, dans les temps impartis, en illustrant de démonstrations pratiques bien préparées.

Il a répondu à de nombreuses questions des membres de son jury, en défendant son point de vue.

Pour ces raisons le jury a décidé à l'unanimité de lui attribuer le grade de Docteur en Informatique de l'Université Paris8-Vincennes-Saint-Denis, avec la mention très honorable.

Fait à St Denis le 17 décembre 2014



An online reversed French Sign Language dictionary based on
a learning approach for signs classification

Pattern Recognition Letters

2015



Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

An online reversed French Sign Language dictionary based on a learning approach for signs classification[☆]



Mohammed Zbakh, Zehira Haddad^{*}, Jaime Lopez Krahe

THIM, CHART EA4004, Université Paris 8, 2 rue de la liberté, 93526 Saint Denis, France

ARTICLE INFO

Article history:

Available online 16 September 2015

Keywords:

Human machine interface
French Sign Language
Classification algorithms
Learning algorithms

ABSTRACT

Sign Language contains very rich vocabulary units which can vary by a simple change in shape, position or movement. The characteristic of this language, to be practiced in space, has encouraged us to adopt a pragmatic approach, which facilitates access to its vocabulary. In this paper, we take advantage of digital techniques to build a human machine interface to create a dictionary of French Sign Language (FSL) to French as a web application. We notice that the FSL has not yet any fixed set of rules to learn this language. We hope that this application could contribute both in spreading and stabilizing the language. To this end, we develop a scalable research system, able to find the meaning of a FSL sign from some features of the sign itself. In this work, we set up a web platform for signs and we analyzed user requests that are connected to this platform. This analysis based on some experiments; aim to identify the parameters necessary to develop an application able to easily find the meaning of the sign in French. Therefore, by studying user behavior, we build a cognitive system that meets to needs.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The language is the main instrument of communication between individuals. This is a code that ensures a level of exchange between the transmitter and the receiver and vice-versa. Thus, it is based on various modalities (speech, writing, or gestures). Sign language (SL) is a language mainly used by deaf people and people with whom they communicate. Similarly, it is equivalent to any other language, as demonstrated by the work of Stokoe of the American Sign Language (ASL) [1]. It would be nice to meet a deaf person and exchange knowledge with it like we exchange with a tourist who speaks a different language. Unlike spoken language, SL is based on gestural-visual channel and non on vocal sound [2]. SL as spoken language has a vocabulary and a grammar. It also has the particular feature of being formed in space. This specificity creates certain difficulties for the computer implementation of this language. It is only recently, now that the computers have gained considerable power of information processing that the computer modeling of this language has become possible. This modeling has been treated with various aspects in the research domain.

Several works treated the sign language recognition by using computer vision. We cite [3] which use deterministic robust features based on hand trajectories in order to propose a sign level classifier. The classifier is based on sequential pattern trees approach by using Kinect features. Another work proposes a machine learning model for Indian sign language recognition using webcam images [4]. This model groups three methodologies: image processing, computer vision and neural networks. The sign interpretation is done by Haar Cascade classifier. In [5], the authors work on the recognition of the 36 Bangali alphabets in real-time using K-nearest neighbors (KNN) classifier. Whereas, other researches treat: automatic SL video processing [6], recognition of SL gestures [7], support for the SL acquisition and learning [8], and the categorization of SL [9]. Most of these approaches [3–7] are basically founded on image processing in order to recognize signs in real-time. While we develop in this work, a signs search engine based on an encoding of the FSL signs and not based on image processing. Thus, we avoid all the constraints related to image acquisition (quality, environmental management, equipment, etc.). Moreover, we do not work on a real-time translation but on a reversed dictionary which is always available. Furthermore, the majority of these approaches [4, 6–9] have treated the grammar of FSL or the conversion of FSL in the form of written language. Our approach that treats vocabularies units does not follow the same approach or the same direction of these researches. In fact, it focuses on the production of a reversed dictionary which does the translation from the FSL to French. In French Sign Language, there exist other works on dictionaries from French to FSL [10] but no previous works dealing

[☆] This paper has been recommended for acceptance by Lledó Museros.

^{*} Corresponding author. Tel.: +33 149407347; fax: +33 149407348.

E-mail addresses: mohammed.zbakh@univ-paris8.fr, zhaddad-bousseksou@univ-paris8.fr (M. Zbakh), zehira.haddad-bousseksou@univ-paris8.fr (Z. Haddad), jlk@univ-paris8.fr (J. Lopez Krahe).

with reversed dictionary (FSL to French) or works based on other techniques for the same production.

This work describes the first initiative of an online reversed dictionary that allows a search in FSL vocabulary. The idea is to provide a simple tool to help users who want to find the meaning of a sign. Lexical dictionaries are considered to be practical gateways between languages. In recent decades, they have been rapidly adapted to new technologies. So much indeed that, they have now for a great part abandoned their classic look of books, to conquer the world of the Internet. This development has allowed them to be more accessible and responsive through the use of different systems of indexing and adequate classification. The main goal of this work is to propose an online dictionary able to help the users to find the significance in French of a sign by inserting some choices of sign parameters. In this regard, we have studied the specification of FSL and created a database using a validation process by FSL experts. Then, we have put our dictionary online and performed a learning approach so that the dictionary can be as effective for non-experts as for FSL experts. In order to achieve this task, we defined two main steps:

- Modeling the vocabulary of the FSL in order to be ergonomically exploitable within the constraints of human-machine interfaces: Based on some parameters of the FSL, we designed a simple model, which does not include necessarily all the strict rules of the sign description.
- Proposing an automatic classification protocol of FSL based on learning algorithms. The learning process will allow us to access to the sign with a minimum of parameters. The aim is to develop a platform capable to take account the user experiments in order to obtain an effective result.

Given the absence of an institutionalized written of SL, we find ourselves confronted with the difficulty of analyzing this language. We must also respect the ergonomic constraints of human-machine interfaces. Regarding this, the main objective of this research is to find a compromise between the modeling of signs and consistency of results. To do this, the developed tool must therefore answer the needs of users who can be in our case: hearing people in the learning phase who need to be familiar with the FSL lexicon; and deaf people, for whom the FSL is the first language and who seek to know the French translation of a sign.

During this research, we worked with a set of FSL parameters, which may present a satisfactory representation space of a sign. The user gives the parameters of an unknown sign and the system brings him a tag cloud that answers to his request. For this purpose, the developed system behind the interface computes similarity values and follows classification and learning algorithms in order to recognize the sign and assign to it the French signification.

This paper is organized as follows. In Section 2, we present the used modeling sign parameters for the developed system. Sections 3 and 4 are devoted to the experiments. In these sections, we describe the learning approaches and the classification algorithms. We present and analyze the obtained results. The last section concludes the work.

2. Modeling parameters and encoding system

Recall that SL is practiced in spatial space. This special character poses some difficulties related to the computer implementation of this language [6, 9]. There are two main difficulties: the first is that we do not have good defined rules concerning the FSL parameters (for example: the number of handshapes). The second difficulty concerns the procedure of converting easily the parameters to an exploitable format without overloading the system. In fact, the real challenge that needs to be met is to know what we must identify and keep from the large amount of information to be processed [11]. This permits to achieve a more flexible communication facilitating the communication between the communities of deaf and hearing people.

Table 1
Number of handshapes regarding different authors.

| Author | Number of handshapes |
|----------|----------------------|
| Cuxac | 39 |
| IVT | 61 |
| Bonnal | 44 |
| Braffort | 55/139 |

Researchers who are interested in the study of the SL structure have different views regarding the number of features used to define a sign. For example, Stokoe selected three parameters for the ASL structure [12]. It is based on 55 symbols grouped into three groups: handshape, position and movement. These parameters have been extended in Battison's work with the addition of the palm orientation of hand as fourth parameter of signs [13]. However, the majority of researchers believe that signs are made up of five parameters including facial expressions [14]. We can find also the concept of the dominant hand and the dominated hand in the SL. The dominant hand, generally the right hand for right-handed people, is one which performs a gesture during the production of a sign. For ergonomics reasons, in order to build a simple and fast search system, we have decided to work with three parameters instead of five. The reason is the following: if we introduce five parameters and for each of them, we choose one element in a set of varieties, it becomes very heavy especially for a novice user. This process takes a lot of time and makes the searching process very hard to accomplish. So, one of our ergonomic reasons is to reduce the number of parameters in order to simplify the process and to increase its speed. The reduced number of parameters is compensated by the use of an intelligent classification which learns from the behavior of users. The three selected parameters are:

2.1. Handshape

The handshape shows the spatial morphology and gestures of the hand. Currently in the FSL, no fixed number of handshapes is officially defined or recognized. We have found a wide variety of proposals. Boutora conducted a study on the categorical perception of manual handshapes in FSL based on the comparison of some previous studies made in the field of ASL [15]. The FSL specialists do not all agree on how to define handshapes. According to Cuxac [2], there are 39 basic handshapes. However, Braffort [7] has set up 55 handshapes from previously defined 139 handshapes. This number was reached by following articulatory measures of hand fingers in the course of making a gesture recognition model.

Table 1 shows that there is no consensus between FSL experts on the exact number of handshapes. There are also no rules to define it. At the first, when we started the encoding process, we tried to encode the signs with 39 handshapes [2]. Next, we faced to the first obstacle: some signs have handshapes which are not present in the 39. We concluded that some details are missing. The others works contain a lot of details [7] which can disturb the users. So we decided to propose our system of handshapes.

After a thorough study of statistics of FSL videos, we selected an initial number of handshapes, based on ergonomic criteria. To avoid a too high workload in our dictionary interfaces, we selected 59 handshapes which seemed sufficient to access the vocabulary of our corpus. These 59 handshapes must be presented in an ergonomic interface which is adapted for HCI (Human Computer Interaction). For this, in order to facilitate this HCI, we propose a handshapes distribution in homogeneous groups of observations. Each group is well differentiated from the others. The goal is to can offer a selection inside each group. However, when codifying a sign, we do not carry out the classes but just the handshape.

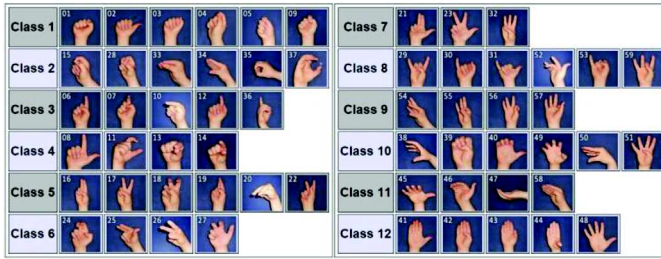


Fig. 1. Classes of handshapes.

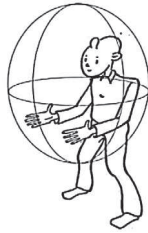


Fig. 2. Signing space.

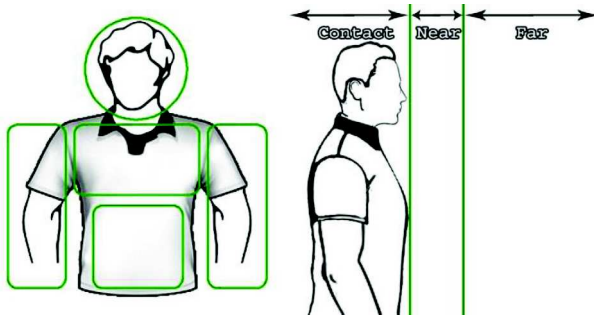


Fig. 3. Positions in our system.

Fig. 1 lists all the 59 used handshapes in our system. We classified them into 12 classes according to the level of similarity between them. For this, we requested the FSL experts to assign a measure of similarity between all the handshapes. This measure value is between 0 and 10 (0 means the same and 10 means very different).

2.2. Position in space

Among the parameters of a sign, we have the position occupied by the hand in space. Fig. 2 shows the 'signing space' that represents the spatial interval in which the signs are signed.

The classical position model in IVT [16] offers a detailed division of the signing space. Their model does not match with our ergonomic constraints, especially for a person that is new to the FSL. It offers 15 positions on the body of the signer, eight positions solely for the face. There are also positions on the neck and the torso also divided into three positions. Add to this, there are positions on the belly and three other positions on the hand. We propose a new model that contains five positions on the frontal plane, merging some positions in order to remove certain ambiguities of interpretation. They are presented on a vertical plane parallel to the torso, and consist of the following: 'head', 'chest', 'abdomen', 'right hand' and 'left hand'. Then we added a criterion representing the distance between the hand and the body. In this regard, we determined three possible positions. The hand may be in contact with the body, near the body, or far from the body. Fig. 3 shows the proposed positions.

Table 2 lists the abbreviations given to the position entity in our database.

Table 2

Structure of the position entity.

| IdPosition | Position |
|------------|-----------------|
| HC | Head-contact |
| HN | Head-near |
| HF | Head-far |
| TC | Torso-contact |
| TN | Torso-near |
| TF | Torso-far |
| AC | Abdomen-contact |
| AN | Abdomen-near |
| AF | Abdomen-far |
| RC | Right-contact |
| RN | Right-near |
| RF | Right-far |
| LC | Left-contact |
| LN | Left-near |
| LF | Left-far |

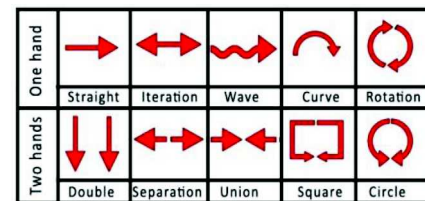


Fig. 4. Movements of the hand.

Table 3

Structure of the movement entity.

| IdMove | Move |
|--------|------------------|
| D1 | Straight |
| C0 | Curve |
| OD | Wave |
| IT | Iteration |
| R1 | Rotation |
| D2 | Double |
| SP | Separation |
| UN | Union |
| R2 | Circle |
| C4 | Square |
| 00 | Without movement |

2.3. Movement

The movement is the action taken by the dominant hand when signation. This is the path followed by the hand in space [17]. Several works, especially in the signs recognition domain use the hand trajectory as in the work of Starner [18]. This criterion represents a greater complexity in the definition of a sign, because the motion affects both the wrist of the hand and/or fingers. In a limited geometric space, movement paths can be interpreted differently between one person and another, especially between two similar paths.

To model the movement of the hand, we considered two types of movements: straight movement and rotational movement. The preliminary analysis of the signs showed us that we needed to describe the movement more accurately in order to discriminate the results. Therefore, we derived four movements from the straight movement: 'Straight', 'Iteration', 'Separation' and 'Union'. Similarly, we derived three other movements from the rotation: 'Rotation', 'Curve' and 'Circle'. Two movements were added later: 'Wave' motion and movement 'Square' (Fig. 4).

Table 3 explains the nominations description of the movement entity in our database.

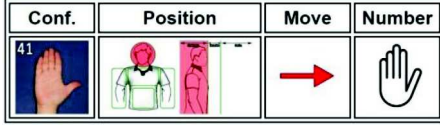


Fig. 5. Graphical representation of sign coding: 'BONJOUR'.

2.4. Final representation

The selected database contains 500 signs chosen from the most used and searched signs. This selection is based on three millions consultations recorded in a previous work [10]. Each sign was modeled by three parameters (handshape, position and movement). We have two types of signs: signs with one hand and signs with two hands. For this reason we add a fourth parameter which corresponds to the number of hands. Each parameter is thus represented by a code: (Conf, Position, Move, Number). Fig. 5 shows a simple representation of sign in the database. For example the sign encoding for 'BONJOUR' is:

([41], [HC], [D1], [01]).

When the sign is stored in the database, we can generate the corresponding encoding representation of its criteria. This representation will facilitate the comparison of our reference encoding with different encodings saved during the experiments.

Finally, when the user searches the meaning of a sign, he selects one element for each requested parameter in order to translate his perception of the sign. In the first time, the algorithm compares the user encoding with the sign encoding (the reference in the database) by calculating the dissimilarity between the two encodings and displays the results like a tag cloud. In the next section, we will describe how to exploit the user's requests in order to analyze them and to find the best approach to give the good result.

3. Experiments based on measuring of dissimilarities

The goal of this first part of the experimentation is to exploit the dissimilarity matrices to calculate the dissemblance between the parameters modeling the signs and those corresponding to the subjective perception of the users. In this section, we introduce the part of the experiments. We explain our approach for measuring dissimilarities. Then, we present and discuss the obtained results. Finally, we propose our learning model based on this measure.

In this experiment, we recorded 6 775 observations by 36 participants. The conditions of this experiment are:

- The level of language and age of participants varied.
- Signs are generated randomly.

During the experiment, participants followed a complete search pattern. Besides, the method for calculating the dissimilarity was adjusted according to the feedback of participants. Fig. 6 shows a general diagram of the experimental procedure. After having logged to the platform, the user receives a video sign to code it. He selects the parameters and submits them. In this phase, these parameters form an object of user encoding which will be transmitted to the application. This last seeks the dissimilarity matrices to calculate the global dissimilarity between the signs and the user encoding. In the next step, the obtained results are presented and sorted in descending order of similarity, in order to give the French significance of the generated sign. Finally, the participant validates the searched result by checking its encoding. He can also check the relevant reference to the sign code by clicking on the result.

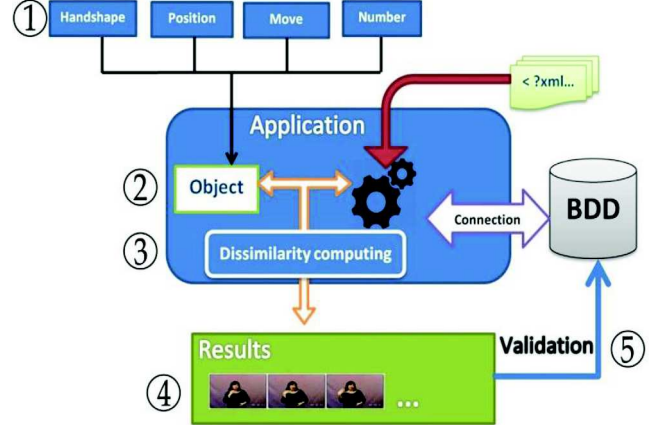


Fig. 6. Diagram of the search procedure. The different steps are: 1. Selection of parameters. 2. Object creation. 3. Dissimilarity computation. 4. Displaying the results. 5. Encoding validation.

3.1. Approach for measuring dissimilarities

First, we have initialized three matrices for the three parameters by FSL experts. These matrices contain the dissimilarity degrees of each sign parameter. The values are then converted in degree of dissimilarity.

In the second stage, we determined the dissimilarity function. This function links the participants encoding and the reference encoding in order to compute the global dissimilarity. It is based on the following matrices values:

$$\begin{aligned} dis(S_{ref}, S_{user}) = & \sum_{i=1}^2 \alpha (dis(C_{ref(i)}, C_{user(i)})) \\ & + \beta (dis(P_{ref(i)}, P_{user(i)})) \\ & + \gamma (dis(M_{ref(i)}, M_{user(i)})) \\ & + \delta (dis(N_{ref(i)}, N_{user(i)})) \end{aligned} \quad (1)$$

where:

$dis(S_{ref}, S_{user})$: The global dissimilarity between the sign and the encoding of the user.

$dis(C_{ref(i)}, C_{user(i)})$: The dissimilarity between the handshapes.

$dis(P_{ref(i)}, P_{user(i)})$: The dissimilarity between the positions.

$dis(M_{ref(i)}, M_{user(i)})$: The dissimilarity between movements.

$dis(N_{ref(i)}, N_{user(i)})$: The dissimilarity between the hands number.

$\alpha, \beta, \gamma, \delta$: The learning coefficients with $\alpha + \beta + \gamma + \delta = 1$.

$I = 1, 2$: The number of configurations in a sign.

Once the dissimilarity function defined, we focus on the calculation of the learning coefficients. First, we created the confusion matrix signs according to the following values:

$$\alpha = \beta = \gamma = \delta \quad (2)$$

The measure of dissimilarity between all X signs gives a matrix $M[i, j]$ ($1 \leq i, j \leq X$) symmetric and zero diagonal. However, if ever $M[i, j] = 0$ when $i \neq j$, this proves that there is much confusion between the two signs i and j in the representation space following a surjection. This can be explained by a similar encoding signs despite their visual difference. To avoid this problem, we need either to expand the representation space with other parameters which implies heavier requests; or to accept the existence of homosigns.

The last important step is to adapt the learning coefficients to the different results. We re-run the scenarios of the participants, and tested them with different learning coefficients (α, β, γ , and δ). This



Fig. 7. Success rate of the handshapes.



Fig. 8. Monitoring window for the handshape '01'.

process has significantly reduced the effort and time in this experiment. During this stage, we adjusted the values of the coefficients to extract the optimal dissimilarity function. To better understand the results, we conducted further analysis on the state of the parameters in this experiment. We studied the success rate of all the parameters, the percentage of the distribution and the factors that cause the selection ambiguity.

3.2. Results of measuring dissimilarities

In this section, we present the obtained results of the previous experiment for each parameter and discuss these results.

• Handshape

Fig. 7 shows the success rates of each handshape during this experiment. We note that some handshapes present no problem of ambiguity for participants like: '10', '21', '29', '36' and '56'. We have also handshapes whose the success rate is very high like: '02', '08' and '31'. On the other hand, we have handshapes whose the success rate is very low, such as: '03', '05', '12', '30', '44' and '50'. These handshapes are influenced by others handshapes belonging the same classes, which dominate the majority of selections.

We developed a monitoring system to trace the different states of each handshape in all observations in order to study the behavior of each handshape in this experiment. Fig. 8 shows an example of this monitoring. We setup '01' which has the form of a closed fist. This handshape is shown in the reference signs 28 occurrences, and 71 occurrences in the participants encoding. At the bottom of the figure, we have the list of handshapes selected instead of the handshape '01'. We have: handshapes with uncolored backgrounds which correspond to handshapes of the same class '01' and handshapes with a black background '35', '41', '44' and '55' which correspond to handshapes of others classes. From this figure, we can see the details of each selection by clicking on the image of the handshape. We note that this handshape '01' has been selected exactly 53 times out of 71

Table 4

Confusion matrix for all handshapes classes and all attributes. Class 1 has 13 examples wrongly predicted as class 2.

| Class | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|-------|-----|-----|-----|-----|-----|----|----|-----|----|-----|----|-----|
| 1 | 264 | 13 | 0 | 2 | 0 | 1 | 0 | 0 | 1 | 8 | 1 | 2 |
| 2 | 2 | 192 | 0 | 19 | 1 | 0 | 1 | 2 | 1 | 17 | 5 | 7 |
| 3 | 0 | 0 | 189 | 19 | 2 | 2 | 2 | 3 | 0 | 0 | 2 | 0 |
| 4 | 0 | 0 | 3 | 108 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 1 | 1 | 0 | 1 | 177 | 14 | 4 | 0 | 5 | 6 | 0 | 1 |
| 6 | 0 | 2 | 0 | 1 | 13 | 15 | 5 | 0 | 2 | 0 | 0 | 0 |
| 7 | 0 | 2 | 0 | 0 | 12 | 2 | 62 | 0 | 0 | 1 | 1 | 0 |
| 8 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 121 | 2 | 3 | 5 | 0 |
| 9 | 0 | 1 | 1 | 8 | 0 | 0 | 0 | 0 | 68 | 2 | 1 | 0 |
| 10 | 3 | 23 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 189 | 5 | 38 |
| 11 | 3 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 61 | 35 |
| 12 | 1 | 10 | 0 | 3 | 6 | 0 | 1 | 0 | 0 | 14 | 33 | 408 |

in total. Note that some anomalies can be found like the handshapes '41', '44' and '55' that do not belong to the same class.

The experimental results of this analysis have enabled us to generate the confusion matrix handshapes (Table 4). The confusion matrix is a specific table layout that allows visualization of the performance and errors. Each column represents the instances in a predicted class, while each row represents the instances in an actual class. So with this matrix, it becomes easy to see when the system is confusing two classes.

• Position

For the hand positions, we proceeded in the same way as the handshape. The results of this model are presented in Fig. 9.

As in the handshapes we investigated each position selected by the participants in the recorded observations. We noticed an influence of hand movement on its position. Some participants take the position after the movement. For example, the sign 'BONJOUR' is carried with the handshape '41' in contact with the head and the hand moves toward the outside of the body. However, some participants select the position 'head-near' and no 'head-far'. Fig. 10 shows the detailed analysis of the position 'head-contact'.

As Fig. 10 suggests, the position 'HC' (head-contact) was confused with almost all other positions. Among 478 selections, positions having a number of less than or equal to 10 selections were ignored in

| Positions | | |
|--------------|--------------|--------------|
| | | |
| 'AC' 33.33 % | 'AF' 16.67 % | 'AN' 18.75 % |
| | | |
| 'HC' 60.73 % | 'HF' 0 % | 'HN' 41.74 % |
| | | |
| 'LC' 35.71 % | 'LF' 0 % | 'LN' 11.11 % |
| | | |
| 'RC' 17.24 % | 'RF' 25 % | 'RN' 32.78 % |
| | | |
| 'AC' 60.82 % | 'AF' 33.33 % | 'AN' 64.15 % |

Fig. 9. Success rate of the positions.

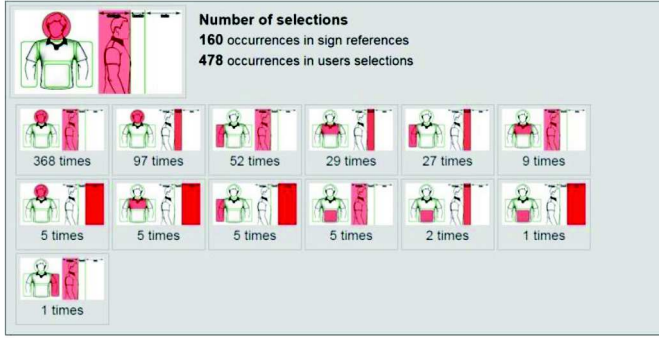


Fig. 10. Monitoring window for the position head-contact.

the calculations. By analyzing selections, we have a success rate of 76.9% (368 on 478 occurrences). The position that causes the highest ambiguity is the position 'HN' (head-near) with a value of 97 on 478 times. In geometric space, both these positions are very close, as the front position is the same: 'Head' and the main difference remains in the horizontal position.

Table 5 presents the confusion matrix of hand positions relative to the body. Maximum selection values are shown in bold. Among all the positions, the positions 'AN', 'AC', 'RC' and 'RN' were not associated with a correct selection during the experiment. Positions 'HF', 'AF', 'RF', 'LN' and 'LF' are represented by very low values. These are the locations that designate the position (F = Far). This can be interpreted as the difficulty to perceive the distance from the hand to the body.

• Movement

In Fig. 11, we have very varying percentages between 45.6 and 81.8%. As previously mentioned, we have two main movements: right and rotation. Other movements are derived from these two types of movements. This explains the variety of values. Participants easily confused between a straight horizontal movement and a movement curve on the same plane.

Detailed analysis showed that there are other factors that affect the status of this parameter. Fig. 12 presents the window of the detailed analysis for the movement 'Curve'. In this figure, we have 263 selections of this type of movement ignoring the movements which have values less than 4 occurrences. There are three types of movements that have been confused with the movement 'Curve': 'Straight', 'Rotation' and 'Without movement'.

Regarding the type of movement 'Straight', we have 21 selections among 263 (7.9%). We noticed that the two movements are very close for certain signs. The signs that have a horizontal movement are less detected than the signs that have a vertical or frontal movement. For

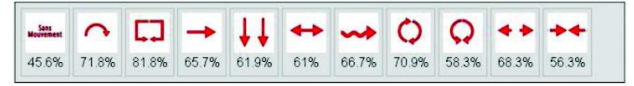


Fig. 11. Success rate of the movements.

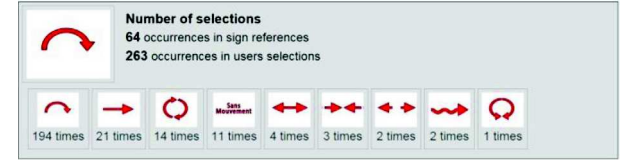


Fig. 12. Monitoring window for the movement 'curve'.

example, the sign 'Donner' that is performed by the handshape '33' with movement 'Curve' to the outside of the body, has been confused with another movement 'Straight' to the outside of the body in some observations.

Table 6 shows the confusion matrix of hand movements. We notice that the maximum values are concentrated to the diagonal of the matrix, but we also have a very large dispersion of values in the types of movement: '00', 'D1', 'IT' and 'UN'.

3.3. Learning model based on the dissimilarity function

This work is based on parameters that correspond to the sign definition. In this regard, the approach is to update the dissimilarity matrices through the experiments. So we worked on the dissimilarity matrices based on confusion matrices. The objective is to adjust the degrees of dissimilarity so that they best meet the global dissimilarity calculation function. In other words, this approach consists in reducing the dissimilarity between the parameters that have been obtained in a significant number of selections, and to increase the dissimilarity of parameters that received a low number of selections [19]. For this, we worked on a probability matrix based on the confusion matrix to adjust the values of the dissimilarity matrix. This adjustment (update) is controlled by the following criteria:

- The dissimilarity between the parameter itself remains unchangeable.
- The dissimilarity receives the maximum value of 1 if the number of selections associated is 0.
- The new dissimilarity value is calculated as:

$$\text{new_dis}(C_i, C_j) = \text{old_dis}(C_i, C_j) * \text{probability}(C_j) \quad (3)$$

If the probability of the parameter which change is greater than the probability of the parameter of reference (the number of

Table 5
Confusion matrix for all positions and all attributes.

| Ref | HC | HN | HF | TC | TN | TF | AF | AN | AC | RC | RN | RF | LC | LN | LF |
|-----|------------|-----------|----|-----------|------------|----|----|-----------|----|----|-----|----|-----------|----|----|
| HC | 371 | 98 | 5 | 9 | 29 | 5 | 1 | 2 | 5 | 52 | 27 | 5 | 1 | 0 | 0 |
| HN | 28 | 92 | 9 | 4 | 53 | 5 | 0 | 0 | 0 | 4 | 20 | 3 | 1 | 0 | 0 |
| HF | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TC | 1 | 0 | 0 | 59 | 28 | 2 | 0 | 1 | 0 | 5 | 1 | 1 | 1 | 0 | 0 |
| TN | 14 | 82 | 3 | 13 | 546 | 34 | 5 | 15 | 2 | 14 | 106 | 6 | 3 | 5 | 2 |
| TF | 0 | 1 | 0 | 2 | 14 | 13 | 2 | 2 | 1 | 0 | 1 | 3 | 0 | 0 | 0 |
| AF | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AN | 1 | 2 | 0 | 4 | 43 | 2 | 3 | 15 | 1 | 2 | 5 | 2 | 0 | 0 | 0 |
| AC | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 13 | 7 | 0 | 0 | 0 | 0 | 0 | 0 |
| RC | 1 | 9 | 0 | 3 | 5 | 0 | 0 | 2 | 0 | 5 | 3 | 0 | 0 | 1 | 0 |
| RN | 3 | 22 | 2 | 3 | 90 | 7 | 3 | 14 | 0 | 7 | 81 | 9 | 0 | 1 | 1 |
| RF | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 0 |
| LC | 1 | 0 | 0 | 6 | 4 | 0 | 0 | 1 | 0 | 4 | 2 | 0 | 15 | 7 | 2 |
| LN | 0 | 0 | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 1 | 0 |
| LF | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 3 | 1 | 0 |

Table 6
Confusion matrix for all movements and all attributes.

| Ref | 00 | C0 | C4 | D1 | D2 | IT | OD | R1 | R2 | SP | UN |
|-----|-----|-----|----|-----|----|-----|----|----|----|-----|----|
| 00 | 256 | 54 | 13 | 141 | 35 | 124 | 25 | 31 | 11 | 23 | 26 |
| C0 | 18 | 110 | 0 | 30 | 8 | 17 | 2 | 16 | 2 | 9 | 3 |
| C4 | 3 | 0 | 3 | 2 | 3 | 1 | 2 | 0 | 2 | 1 | 0 |
| D1 | 48 | 27 | 2 | 235 | 14 | 22 | 13 | 20 | 4 | 18 | 8 |
| D2 | 14 | 7 | 2 | 11 | 65 | 10 | 6 | 8 | 4 | 0 | 5 |
| IT | 39 | 15 | 7 | 36 | 7 | 170 | 30 | 12 | 3 | 7 | 18 |
| OD | 2 | 4 | 0 | 5 | 2 | 6 | 27 | 3 | 0 | 0 | 0 |
| R1 | 18 | 12 | 0 | 5 | 8 | 8 | 1 | 90 | 4 | 3 | 4 |
| R2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 8 | 3 | 2 |
| SP | 16 | 3 | 1 | 12 | 13 | 7 | 1 | 16 | 13 | 101 | 16 |
| UN | 13 | 4 | 2 | 17 | 12 | 11 | 1 | 12 | 6 | 8 | 46 |

selections C_j is greater than the number of selections C_i), we assign a value of 0 to indicate the dissimilarity between the two parameters.

Once the dissimilarity matrices uploaded, we have worked on learning coefficients in the weighted function to calculate the dissimilarity. Given the material constraints and the large amount of data to process, we took a sample of 600 observations stored in this experiment. We applied the weighted function to calculate the dissimilarity based on all observations. Note that the coefficients will be updated after each execution for each observation and that we saved the number of good signs in each execution and the five first results of each execution. The choice of the five first results is performed by the presence of homosigns phenomena. In this case, the order in the displaying results is based on the frequency of visits.

Table 7 shows a sample of the first results of the implementation in weighted function sorted in descending order of the total number of results.

In the table, from left to right, we have the four values of the used coefficients. Then there are five columns in the middle to indicate the percentage of correct answers for each function in the five first positions, and in right columns, we have the total percentage of good observations in 600 samples. The first observation indicates that the maximum of good answers is positioned in the first rank. This means that the parameters selected from these experiments are sufficiently able to give good results. Among the functions, four cases are given a coefficient greater than 0.4 for handshake. We can conclude that the weight of the handshake greatly affects the results of the calculation function. Furthermore, the position of the hand and its movement has the same degree of influence with a small advantage of the hand movement in certain signs.

The evaluation of the final results in this experiment allowed us to propose a strategy of self-correction for the signs. This strategy is mainly based on user feedback. Indeed, we found that a sign does not necessarily have a single encoding, but can have multiple encodings. For this, we consider the encoding of the user in the case where there is a systematic repetition of this encoding for this sign.

4. Experiments based on users behaviors

The variation in results among different users of our system prompted us to improve the performance of our algorithm for more

precision in the query results. To do this, each user behavior was treated with taking account their eventual errors.

In the following, we will explain a new exploring approach of the used experimental data. This approach is based on the level of precision in previous experiments performed by users. First, we analyzed the information corresponding to each encoding by establishing it's a network of related signs. Next, we applied a learning based on the stored signs encodings.

We observe that one of the most common mistakes is caused by the ignorance of one of the handshapes in signs with more than one handshape. Thus we found that the user tend to minimize the sign description by taking into account just the last handshape.

For example, the sign 'BLANC' is a sign consisting of two handshapes as shown in the following figure:

The closed hand accolade neck signer that changes its shape to an open hand in the sign. Fig. 14 shows the encoding of this sign in our database. We have the handshake number '35' which is transformed into a handshake '38' by a straight movement outwardly of the body.

So, in the case of sign with two configurations, we compare the first configuration with the first and the second configuration with the second. If the user ignores the first configuration and enters just the second, the system compares the second configuration with the first. The following example illustrates this problem:

The sign 'BLANC' is coded as follows:

[38-HN-00-01-35-HC-D1-01]

If the user says:

[35-HC-D1-01]

The result will be the comparison:

[38-HN-00-01] AND [35-HC-D1-01]

In this case, the dissimilarity may be considerable, which explained a source of errors. These errors are more common among apprentices of the FSL as among experts. Nonetheless, we consider that future users of this application will be much more often apprentices than experts, and the system can be programed to adapt itself to the user level. The following sections propose the solution of this problematic. The basic idea is to exploit the errors of users in a learning system in order to enhance the results for the novice FSL users if they make some mistakes in their choice of parameters. For this, we present in first, the learning approach based on user profiles, followed by the proposed leaning phase experiments. After this, we present the three types of learning: on the entire table, on the half database and the cross learning. We conclude with the classification and the obtained results.

4.1. Leaning approach based on user profiles

In this experiment, we have two datasets. The first is a dataset of signs composed by the encoding of the parameters (the reference dataset). The second is a dataset of experiments. This dataset saves all the encodings of users.

Table 7
Table sorted results weighted functions.

| C | P | M | N | 1st | 2nd | 3rd | 4th | 5th | 1+2 | 1+2+3 | 1+2+3+4 | Total |
|------|------|------|------|--------|--------|-------|-------|-------|--------|--------|---------|--------|
| 0,41 | 0,27 | 0,23 | 0,09 | 37,83% | 11,53% | 7,06% | 4,35% | 3,29% | 49,37% | 56,42% | 60,77% | 64,06% |
| 0,45 | 0,25 | 0,25 | 0,05 | 38,27% | 11,56% | 6,67% | 4,52% | 2,97% | 49,83% | 56,50% | 61,03% | 63,99% |
| 0,4 | 0,2 | 0,32 | 0,08 | 38,38% | 11,44% | 6,90% | 3,91% | 2,90% | 49,82% | 56,71% | 60,63% | 63,53% |
| 0,4 | 0,25 | 0,3 | 0,05 | 38,80% | 11,21% | 6,51% | 3,53% | 2,87% | 50,01% | 56,52% | 60,05% | 62,92% |
| 0,37 | 0,26 | 0,3 | 0,07 | 38,94% | 10,81% | 6,69% | 3,72% | 2,97% | 49,75% | 56,44% | 60,16% | 63,13% |

Table 8
Distribution of encoding [41-TN-SP-02] in the database.

| Signs | Encoding | Experiments |
|------------------|---------------|-------------------|
| | [41-TN-SP-02] | Jour (18); |
| Franç (Honnête); | | Franç (14); |
| Jour; | | Simple (6) |
| Livre; | | Livre (4); |
| Nager; | | Nager (4); |
| Natation; | | Natation (4); |
| Sauvage | | Sauvage (4); |
| | | Tant pis (4); |
| | | Bibliothèque (2); |
| | | Divorce (2); |
| | | Verglas (1); |
| | | Vierge Marie (1) |

We have 2,500 videos of FSL signs. In order to facilitate the treatment, we choose a sample of 500 signs to be coded by FSL experts. We started the experiment with a subset dictionary test (T) of these 500 referenced signs (videos) that were used in the experiment. Once the proposed method validated, we generalize it on the whole of the database.

During our analysis of T, we remarked some homosigns (signs that have the same appearance with different significations), for example: 'MERCI' and 'BONJOUR'. These signs will have certainly the same encoding. Therefore, T corresponds to a total of 464 different encodings. Thus, the number of referenced encodings $Ref = 464$.

In our experiment, we present to the user a sign (a video) of T and we ask him to code it. The user codes the presented sign based on its corresponding video, without knowing the referenced encoding which is initialized by FSL experts. So, we collect a database of 6,775 observations performed by 36 participants, on which we apply the learning process. Each observation corresponds to a sign and a user encoding. In this phase, we analyze the experimental data of these observations, and for each sign (video), we examine the behavior according to the assigned encodings.

In the experiment, we have collected 6 775 observations on the 500 presented signs. It is clear that the same sign was encoded by several users. This experimentation produced 2 501 different encodings. Thus, the number of experimental encodings $Exp = 2501$. In total, the number of different encodings Tot in our database is the union of the two sets Ref and Exp.

$$Tot = Ref \cup Exp = 2531 \quad (4)$$

This number does not correspond to the mathematical sum because both Exp and Ref sets have certain encodings in common.

To explain this mechanism, we take the example of the encoding: [41-TN-SP-02]. It corresponds to a handshape which is flat and carried on the torso by two hands in separating movement. Table 8 shows the distribution of this encoding in the database. This encoding is present in six signs of referenced signs by the FSL experts (1st column of Table 8). In contrast, with the experimentation, we recorded 64 appearances of this encoding on 12 different signs (3rd column of Table 8). We notice that there are two signs that have a high frequency with this encoding: (18 occurrences) 'JOUR' and (14 occurrences) 'FRANC'. This indicates that we have a strong possibility that the given encoding matches one of these two signs.

Fig. 15 shows a cloud presentation of signs generated by the encoding [41-TN-SP-02] and based on the frequency of each sign. Signs 'JOUR' and 'FRANC' are represented with a very large font size compared to other signs. This process is performed for all database encodings to determine a global table which will be used in learning of signs (Table 9).

4.2. Proposed learning phase experiments

We can formalize our experiment by this application:

$$h : T \rightarrow \Omega$$

It relies to each video corresponding to a sign of T, an element Ω selected by the user during the experiment. This application allows us to establish a statistical learning table of all signs of T. This represents, for each sign, a set corresponding to a collection of all encodings assigned to this sign during the experiment.

So, we can build a table Ω^* which contains all Ω elements. In each element Ω of Ω^* we have:

- Either one sign with n occurrences.
- Either i signs with their ni occurrences. These signs will be sorted by their occurrences.
- Either no occurrence (nobody assigned a sign to this reference).

Table 9 shows an extract from the global statistical learning table Ω^* . The first column indicates all encodings presented in the database. This is followed by a sequence of pairs of columns. Each pair of columns contains the ID code of the referenced sign and the number of occurrences of this sign for the given encoding.

In following, we propose an approach that is based on this table in order to establish a fast addressing of encodings to referenced signs.

Table Ω^* can be dynamic and enriched by usage. We applied learning process in the table Ω^* according to three protocols:

- Learning on the entire table.
- Learning on the half database.
- Cross-learning.

Each person has a way of seeing a sign and therefore, a way of encoding it. We are facing of a non-deterministic problem since the same sign may be encoded in different ways (different encodings). To address this problem, we use a learning method which is strictly based on experimentation. The system learns the behavior of the users, it learns how they understand and encode a sign. In this case, even if they make mistakes, if they do not encode as the FSL experts, the system learns their mistakes. Thereafter, the system is able to generalize what it has been learned. This is performed in order to assign to an input encoding, the most probable sign that have been previously learned with the encoding in the experimental phase. This process is based on Table 9.

4.3. Learning on the entire table

In this case we have taken all the elements of the table Ω^* as input elements and we compared them with the outputs from the same table. The training set is itself the test database.

Table 10.1 shows the result of learning on all signs of the database. We have 6 775 observations either with one or two configurations which give 8 757 encodings. Note that this number contains multiple occurrences of the same encoding. We saved the number of responses for each ranking (the number of responses located in first class, the number of responses in the second class, and so on). Then, we calculated the percentage of each class by adding the number of previous responses of classes. In Table 10.1, we have 61.21% of correct answers in class 1. Then 80.27% of correct answers to the sum of both classes 1 and 2 and up 95.99% of correct answers to the sum of the first 5 classes. Results of this learning protocol are very encouraging.

Knowing that there are in the database, unique encodings (encoding with one answer), we reapplied learning on the same database without unique cases. This manipulation can avoid scenarios that may affect our results. Table 10.2 shows the final results of this type of learning. We have 7 496 encodings. We notice that there is a difference in results for the first class (54.68% against 61.21%) but the final result is very close and similar to results obtained with the first learning (95.32%).

Table 9

Extract of the global statistical learning table. Each row presents an encoding followed by its corresponding signs (in blue) with the frequency of this lasts (in red).

| Codage | Signe | NB | Signe | NB | Signe | NB | Signe | NB | Signe | NB | Signe | NB | Signe | NB | Signe | NB | Signe | NB | Signe | NB | Signe | NB | Signe | NB |
|-------------|-------|----|-------|----|-------|----|-------|----|-------|----|-------|----|-------|----|-------|----|-------|----|-------|----|-------|----|-------|----|
| 06-HC-D1-01 | 512 | 9 | 139 | 8 | 228 | 7 | 314 | 7 | 294 | 7 | 181 | 6 | 175 | 5 | 253 | 4 | 330 | 4 | 193 | 3 | 194 | 3 | 59 | 2 |
| 41-TN-UN-02 | 362 | 8 | 538 | 7 | 504 | 7 | 547 | 6 | 361 | 6 | 412 | 5 | 416 | 3 | 306 | 2 | 99 | 2 | 446 | 2 | 23 | 2 | 409 | 1 |
| 41-TN-SP-02 | 242 | 14 | 286 | 10 | 287 | 8 | 457 | 6 | 306 | 4 | 550 | 4 | 446 | 4 | 351 | 4 | 349 | 4 | 99 | 2 | 198 | 2 | 497 | 1 |
| 41-TN-IT-02 | 415 | 7 | 395 | 7 | 277 | 6 | 306 | 5 | 266 | 5 | 538 | 3 | 99 | 3 | 351 | 2 | 412 | 2 | 367 | 1 | 147 | 1 | 446 | 1 |
| 48-TN-00-02 | 157 | 7 | 258 | 4 | 149 | 3 | 275 | 3 | 396 | 2 | 541 | 2 | 121 | 1 | 519 | 1 | 406 | 1 | 18 | 1 | 325 | 1 | 342 | 1 |
| 43-TN-D1-01 | 328 | 4 | 445 | 4 | 202 | 2 | 35 | 2 | 126 | 2 | 161 | 1 | 291 | 1 | 21 | 1 | 327 | 1 | 240 | 1 | 226 | 1 | 39 | 1 |
| 41-TN-D1-01 | 23 | 1 | 327 | 1 | 108 | 1 | 21 | 1 | 332 | 1 | 544 | 1 | 252 | 1 | 504 | 1 | 45 | 1 | 335 | 1 | 446 | 1 | 226 | 1 |
| 02-TC-D1-01 | 471 | 21 | 262 | 1 | 492 | 1 | | | | | | | | | | | | | | | | | | |
| 02-HC-D1-01 | 249 | 13 | 146 | 11 | 348 | 8 | 555 | 2 | 185 | 1 | 140 | 1 | 151 | 1 | 105 | 1 | | | | | | | | |
| 41-HC-C0-01 | 279 | 13 | 447 | 3 | 551 | 2 | 504 | 1 | 153 | 1 | | | | | | | | | | | | | | |
| 38-TN-SP-02 | 404 | 12 | 405 | 10 | 154 | 2 | 251 | 2 | 91 | 2 | 121 | 2 | 119 | 1 | 145 | 1 | 177 | 1 | 540 | 1 | 69 | 1 | | |
| 19-HC-IT-01 | 440 | 12 | 425 | 9 | | | | | | | | | | | | | | | | | | | | |
| 18-HN-R1-01 | 370 | 12 | | | | | | | | | | | | | | | | | | | | | | |
| 08-TN-00-01 | 32 | 11 | 255 | 7 | 408 | 6 | 35 | 5 | 33 | 4 | 386 | 1 | | | | | | | | | | | | |
| 08-RN-IT-01 | 303 | 11 | 307 | 10 | 37 | 8 | 408 | 1 | 35 | 1 | | | | | | | | | | | | | | |
| 41-HC-D1-01 | 108 | 11 | 332 | 6 | 153 | 5 | 96 | 4 | 551 | 3 | 339 | 1 | 384 | 1 | | | | | | | | | | |
| 01-RN-IT-01 | 459 | 11 | 443 | 9 | 118 | 2 | 333 | 1 | 246 | 1 | | | | | | | | | | | | | | |
| 49-TN-D2-02 | 517 | 11 | 507 | 2 | 540 | 1 | 69 | 1 | 400 | 1 | 416 | 1 | | | | | | | | | | | | |

Table 10

Results of different learning methods.

| 1.Learning on the entire database | | | | | |
|--|-------|-------|-------|-------|-------|
| Rank | 1 | 2 | 3 | 4 | 5 |
| Number | 5360 | 1669 | 748 | 398 | 231 |
| Percent | 61,21 | 19,06 | 8,54 | 4,54 | 2,64 |
| | 61,21 | 80,27 | 88,81 | 93,35 | 95,99 |
| 2.Learning on the database without unique event of encodings | | | | | |
| Rank | 1 | 2 | 3 | 4 | 5 |
| Number | 4099 | 1669 | 748 | 398 | 231 |
| Percent | 54,68 | 22,27 | 9,98 | 5,31 | 3,08 |
| | 54,68 | 76,95 | 86,93 | 92,24 | 95,32 |
| 3.Learning on test database | | | | | |
| Rank | 1 | 2 | 3 | 4 | 5 |
| Number | 2696 | 541 | 240 | 119 | 59 |
| Percent | 72,03 | 14,45 | 6,41 | 3,18 | 1,58 |
| | 72,03 | 86,48 | 92,89 | 96,07 | 97,65 |
| 4.Cross-Learning | | | | | |
| Rank | 1 | 2 | 3 | 4 | 5 |
| Number | 2639 | 583 | 233 | 123 | 75 |
| Percent | 70,32 | 15,53 | 6,21 | 3,28 | 2 |
| | 70,32 | 85,85 | 92,06 | 95,34 | 97,34 |

4.4. Learning on the half of database

After testing the learning on all the entire of the database, we devise the database into two parts: learning base and test base. During this phase we applied two rules:

- The size of the two bases may be more or less the same.
- All encodings must exist in both databases.

As indicated in Table 10.3, we have 3 743 test encodings with 3 753 learning encodings. We proceeded in the same manner as the first and we got a final score of 97.65% success rate on five classes. This result shows that the learning procedure is really functional although a portion of the database with a small progress of 2%.

4.5. The cross-learning

This protocol provides cross-learning signs in the database. The principle is to consider the training base of the second protocol as a test base and the base that has served as the test base will be the training base. This protocol is used to estimate the reliability of our model and validate the proposed method of learning. Table 10.4 presents the results of this protocol. We have almost the same final result as the second protocol (97, 34%). We find that the results of the three protocols are very close. Similarly, the success rate is very high. These higher success rates are likely due to the learning which is done in an experimental way on user data corresponding to their profiles.

4.6. Classification

During the sign recognition phase, the user set an encoding (Conf, Position, Move, Number), which corresponds to an address of the table Ω^* . At this address we can have different scenarios:

- No occurrence: The searched sign may correspond to a sign that has not been encoded in the dictionary or to an encoding error. In this case, we memorize it and we apply the process of the dissimilarity measures calculation described above to give the corresponding result.
- Only one sign: This sign must be included as a priority in terms of recognition (we activate also the dissimilarity measure procedure that can bring other possible results to display in the "wall").
- i signs: The results correspond to the sort contained in the corresponding address of the array Ω^* .

4.7. Analysis of the obtained results

One of the unexpected results of this experiment is that the encoding errors of users are very repetitive and systematic. Among these errors, we often find the use of the encoding of latest configuration instead of the first, and also some cases where certain hand-shapes are chosen instead of more subtle handshapes. The difference in results compared to the evaluation of the exact signs encoding, is significant. In fact, if we take into account the exact encoding, we obtain 64% of correct results in the first five signs (Table 7). However,

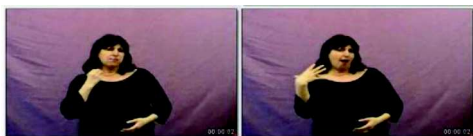


Fig. 13. Graphical representation of coding for the sign 'BLANC'.

| Conf. | Position | Move | Number |
|-------|----------|------------------|--------|
| 35 | | | |
| 38 | | Without Movement | |

Fig. 14. Graphical representation of coding for the sign 'BLANC'.

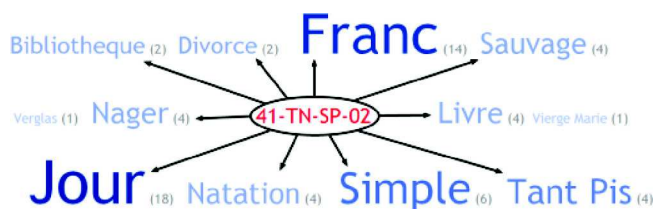


Fig. 15. Frequency of signs for encoding [41-TN-SP-02] in the form of a cloud of words.

| Conf. | Position | Move | Number |
|-------|----------|------|--------|
| 38 | | | |

Fig. 16. Common coding of inexperienced users to sign 'BLANC'.

the use of the proposed original learning approach gives a success rate of 95%. We can add a supplemental advantage to this approach by showing to the users the good encoding, once the user finds the sign sought.

For example we have the example of the sign 'BLANC'. This sign has two handshapes. It starts with a closed hand adjacent to the neck which then moves away outwardly opening fingers. The sign encoding in our database is presented in Figs. 13 and 14. We remark that many users seek the sign 'BLANC' from the following encoding [38-HC-D1-01] (Fig. 16).

The nearest words of this encoding are:

1. COL ROULÉ;
2. TRISTE;
3. SÉRIEUX;
4. **BLANC**;
5. CHAUD;
6. HARICOT BLANC;
7. BALEINE;
8. CRAPAUD.

So with this approach, the user finds the searched word and we indicate him the correct code by showing him the corresponding video in order to he knows encoding. This will not be recorded, for not converge the system to reference encodings and for not biasing the learning mistakes process.

We implement in this work an approach that adapts to users. In the case of sign containing many configurations, the proposed approach allows correcting errors in the confusion of configurations order. These errors are proven common for the non-expert consultants. In case where there are no results to fulfill the "wall", it is still possible to use the method of dissimilarity calculation. Moreover, we showed that it is quite possible to adapt the proposed method to the user according to his profile by storing it and remembering it.

5. Conclusion

The constraints that we have considered in this work, for human-machine interfaces and relevance of the sought results, are respected in the conception and in the final realization. In fact, with simple interfaces and clear questions, the user may happen to find an optimal response in the search of a sign.

In this study, we developed an original system of signs classification as an online reversed dictionary FSL to French. It is the result of a learning cognitive work that permits to find the sign meaning in a dynamic and simple manner.

In this research we observed major inconsistencies between the theoretical approach (the parameters set by the FSL experts) and the reality practiced by users of our system. The experiments show that there is a significant difference between the vision of a user for a sign and the vision of other users. Therefore, we have ensured that our system behave differently between a user practicing the FSL and an others who are in the learning phase of the FSL. This has been done with a compromise between ergonomics interfaces (facility of research) and the quality of results (sped response).

The research sign process in the sense FSL to French is done by a succession of gestural parameters. There are three parameters which are: handshape, hand position and hand movement. A fourth parameter that is the number of hands, is used to play the filter role of results. Our algorithm takes these parameters as an input by preparing them in an appropriate and standardized vector form. After this, the proposed system calculates dissimilarity between user inputs and signs indexed in the database.

In this work, we demonstrate the importance of a system that juggles between different users profiles. We have two methods of calculation: a calculation of parameters and a calculation method by learning. The first method consists of applying a calculation of overall dissimilarity between the parameters entered by the user and the parameters of each sign in the database (determined by FSL experts). This method is more useful for the FSL practitioner where the risk of committing encoding errors is very low. The second method is based on experimental data; thereby it can handle errors done by users in the learning phase. The proposed reversed dictionary was fed by practical observations of the participants, as well as theoretical guidance of specialists.

The obtained results show the efficiency of the proposed approach. The developed dictionary is functional. It is online for free use at the following address:

<http://www2.univ-paris8.fr/ingenierie-cognition/master-handi-di/index.php>

The proposed system is exposed to an operation of 'updating' that permits to recalculate variables and redefine signs after the learning phase. This dictionary which is started with indexed signs, feeds by Internet consultations and observations of users. The vocabulary enrichment is a strategy that we have put in place recently and which will ensure the continuity and the development of our dictionary. We believe that after the enrichment procedure of the dictionary on a long term, representation space will be influenced and it will become more condensed. The current encoding vector (Conf, Position, Move, Number) will generate ambiguous results for some very close signs. Discrimination elements of the representation space must be even

more intelligent and robust. As a first solution, we think that a system with an advanced research should solve this problem and give more precision. Do not forget that we have used just three parameters of a sign with a filtering by the number of hands. There are still others parameters which we can be uses like the hand orientation (with the directions: up, down, right, left, inside and outside). This parameter will can remove more ambiguity to a wide variety of signs. Facial expressions also should play a very important role in this work of discrimination. The current system is very successful on 500 signs; the future works consist to overcome the dictionary limitations of enrichment and expansion at sign. This will be done the day that it will be supported by a functionality able to add more criteria dynamically during an ambiguity state.

Another point to improve is the automatic processing of learning coefficients. The actual procedure involves, the automatic use of each coefficient in a range instead of fixed values depending on the encoding.

Already, in the current system status and after viewing the obtained results, a very interesting application can be proposed concerning the pedagogic context of the FSL learning. For this, we propose for the queries that contain encoding errors, to present a correction that proposes the referenced sign to the input encoding.

References

- [1] W. Stokoe, Sign language structure: an outline of the visual communication systems of the American deaf, in: *Stud. Linguistic Occas. Pap.* 8 (1960).
- [2] C. Cuxac, La Langue des Signes Française: les Voies de l'Iconicité, *Faits Langues* 15–16 (2000) Ophrys.
- [3] E. Ong, H. Cooper, N. Pugeault, R. Bowden, Sign language recognition using sequential pattern trees, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012.
- [4] K. Dabre, S. Dholay, Machine learning model for sign language interpretation using webcam images, in: *Proceedings of the 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, 4–5, 2014, pp. 317–321.
- [5] M.A. Rahaman, M. Jasim, M.H. Ali, M. Hasanuzzaman, Real-time computer vision-based Bengali Sign Language recognition, in: *Proceedings of the 2014 17th International Conference on Computer and Information Technology (ICCIIT)*, 22–23, 2014, pp. 192–197.
- [6] F. Lefebvre-Albaret, P. Dalle, *Analyse de vidéos en langue des signes - méthodes et stratégies*, ORASIS, 2009.
- [7] A. Braffort, *Reconnaissance et compréhension de gestes, application à la langue des signes*, 11, Université Paris, 1996 Thèse de doctorat en informatique.
- [8] M.A. Sallandre, *Acquisition d'une langue des signes comme langue première*, John Benjamins Publishing Company, 2010.
- [9] G. Aznar, *Informatisation d'une forme écrite de la langue des Signes Française*, Université de Toulouse, 2008 Thèse de doctorat en informatique.
- [10] http://www2.univ-paris8.fr/ingenierie-cognition/master-handi/etudiant/projets/site_lsf/dico_lsf/recherche.php (accessed 25.09.15).
- [11] M. Zbakh, Z. Haddad, J. Lopez Krahe, Toward a reversed dictionary of French Sign Language (FSL) on the Web, in: *Proceedings of the 14th International Conference on Computers Helping People with Special Needs (ICCHP 2014)*, July 09–11, 2014, pp. 423–430.
- [12] W.C. Stokoe, D.C. Casterline, C.G. Cronenberg, *Dictionary of American Sign Language on Linguistic Principles*, new ed., Linstok Press, Silver Spring, 1976.
- [13] R. Battison, *Lexical Borrowing in American Sign Language*, Linstok Press, Silver Spring, MD, 1978.
- [14] B. Moody, *La langue des signes – Tome 2: dictionnaire bilingue élémentaire*, International Visual Theatre (I.V.T.), Editions Ellipses, Paris, 1986.
- [15] L. Boutora, *Vers un inventaire ordonné des configurations manuelles de la Langue des Signes Française*, JEP, Dinard, 2006.
- [16] International Visual Theatre, *La langue des Signes* 1, 2 et 3. Edition, 1998.
- [17] C. CUXAC, Iconicité et mouvement des signes en langue des signes française, Le mouvement, in: *Actes de la sixième Ecole d'Eté de l'Association pour la Recherche Cognitive (ARC)*, Paris, 1997, pp. 205–218.
- [18] T. Starner, A. Pentland, Real-time American Sign Language Recognition From Video Using Hidden Markov Models. *Tehnia Report TR-375*, M.I.T Media Laboratory Perceptual Computing Section, 1995.
- [19] J. Fournier, M. Cord, Long-term similarity learning in content-based image retrieval, in: *Proceedings of the IEEE International Conference in Image Processing (ICIP'02)*, Rochester, New-York, USA, 2002.

Article 2

Toward a reversed dictionary of French Sign Language
(FSL) on the Web

ICCHP

2014

Toward a reversed dictionary of French Sign Language (FSL) on the Web

Mohammed Zbakh, Zehira Haddad, Jaime Lopez Krahe

THIM CHART, Université Paris8, 2 rue de la Liberté, 93526 St Denis

{mohammed.zbakh, zhaddad-bousseksou, jlk}@univ-paris8.fr

Abstract. On the web, we can find dictionaries for viewing a sign of French Sign Language (FSL), from a word. However, finding a word from a sign is much more complicated. For this purpose, we propose to design a web application to find the meaning of a FSL sign in the French language from the sign's features. In order to do this, we have developed an intelligent system capable of learning and self-improving by feeding off the information presented to it during its use. We have managed to find a middle ground between the reliability of the results and the ergonomics of Human-Machine Interfaces (HMI).

Keywords. Human Machine Interface, classification algorithms, French Sign Language, Learning algorithm.

1 Introduction

The Sign Languages (SL) are based on the visual-gestural channel, which means, it is the gesture that takes the role the phoneme has in the spoken language [1]. This specificity of SL to be transmitted in a space creates difficulties in its modeling. Research in this field has changed considerably with the advance of technologies such as image processing and the emergence of technical devices such as motion sensors.

Within the existing literature, we can find some works on automatic processing of video sign language [2], [3] and others on LSF signs categorization and classification [4]. Additionally, IVT (International Visual Theatre) proposes a bilingual dictionary divided in three parts [5]: The first describes the evolution of communication in the deaf community, it is particularly focused on linguistic and cultural evolution. The two other parts show a set of 4500 signs classified by theme. The dictionary offers the possibility to search through two indexes: an alphabetical one, corresponding to the French translations of signs, and a second one in which signs are sorted by configuration. Note, however, that the search can be quite tedious if we are looking for the meaning of a specific sign as the dictionary does not offer a quick and easy way to access this sign. On top of that, paper dictionaries have certain limits related to their format.

With this project, we propose to develop a web application in the form of a reversed dictionary: French Sign Language (FSL) \rightarrow French. This free web accessible

application provides a wider visibility to the general public. In fact, sign language (SL) has long suffered from a considerable lack of mediation which has been identified as one of the main reasons of the great instability and heterogeneity of this language. Unfortunately, setting up an accessible diffusion system for SL has been proved difficult. This is mainly due to the absence of a language identification protocol and the poor mediation inside this language. SL are also subject to local and cultural diversity, which often lead to different dialects. The reference dictionaries that we propose to diffuse on the net can provide a wider mediation and thus offer a greater stability.

The main objective of this work is to exploit the advantage computers and web networks offer to build a database that can contribute to the stabilization of the FSL. To do this, we will rely on the FSL dictionary developed within our team:

http://www2.univ-paris8.fr/ingenierie-cognition/masterhandi/etudiant/projets/site_lsf/dico_lsf/recherche.php

Here, we present the design of a web platform that offers a compromise between the quickness of the algorithms and the reliability of the result. This compromise has been carefully crafted to offer a system that is both simple (few fields to fill), fast (in computation and response time) and reliable accurate results). To achieve that, we began our analysis by applying some clustering algorithms such as neural networks, decision trees and support vector machines. The results led us to propose a machine learning system based on certain coefficients.

2 French Sign Language : Parameters and Modeling

The definition of a sign in FSL is based on different gestural and emotional parameters [5]. The sign is characterized by one or more hand shapes, the position of the hand in space, its movement, its orientation and the facial expressions that accompany it. As part of this work, we use a database of 3000 signs indexed and stored in a format usable by a web technology. In a first phase, we limit ourselves to the 500 most used signs in FSL. Then, once the system is implemented, it will be able to expand and learn from the other signs. The first step of this research is to model signs. In order to provide a simple and ergonomic sign modeling, we considered the following parameters:

As the first parameter we choose the hand shape. Note that there isn't a universally used and accepted number of hand shapes, even the identification of each sign has not been standardized [6]. Indeed, scholars don't agree on the matter since their classification of hand shaped ranges from 39 to 139 [3], [5], [7], [8]. Consequently, given the absence of agreement on a specific number of hand shapes used in the FSL [5], we selected 59 hand shapes. In order to present to the user a simple and easily accessible interface, we present the list of hand shapes in a catalog of 12 classes according to their similarity (Figure 1). Each class contains at most 6 configurations. The goal of this organization is to facilitate the access to the signs database.

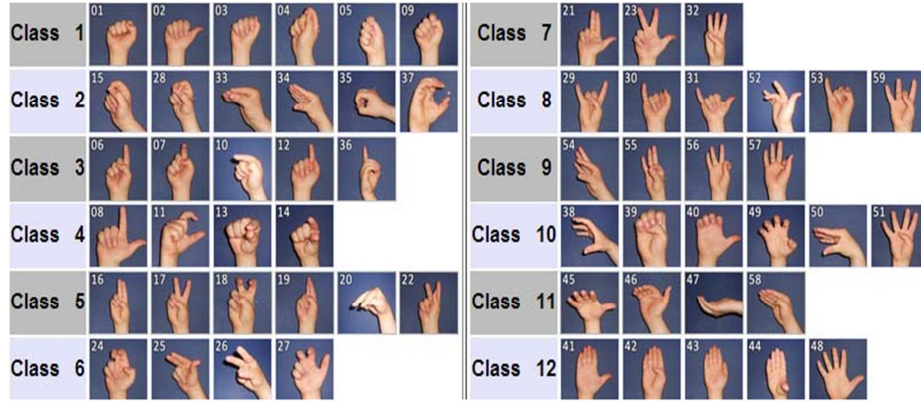


Fig. 1. The 59 hand shapes used.

The second parameter is the hand position. Generally, the proposed positions are quite numerous and complex. Ergonomic constraints guided us towards selecting a simpler model than the model presented in the IVT [5] which cites different detailed positions. We have worked (Figure 2) on five front spaces (Head, Chest, Abdomen, Left arm and Right arm) and three lateral positions (contact with the body, close contact and distant contact). In total, our model manages 15 sites by combining the frontal plane and lateral plane.

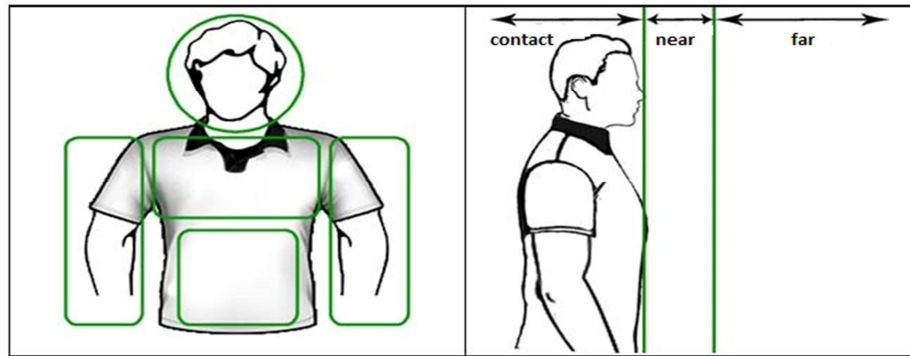


Fig. 2. The different hand positions in space.

Thereafter, we proceeded to the creation of similarity matrices for hand shapes (Figure 3) and the position in space (Figure 4). These matrices are used to initialize our classification system. It was seized by practitioners of the FSL.

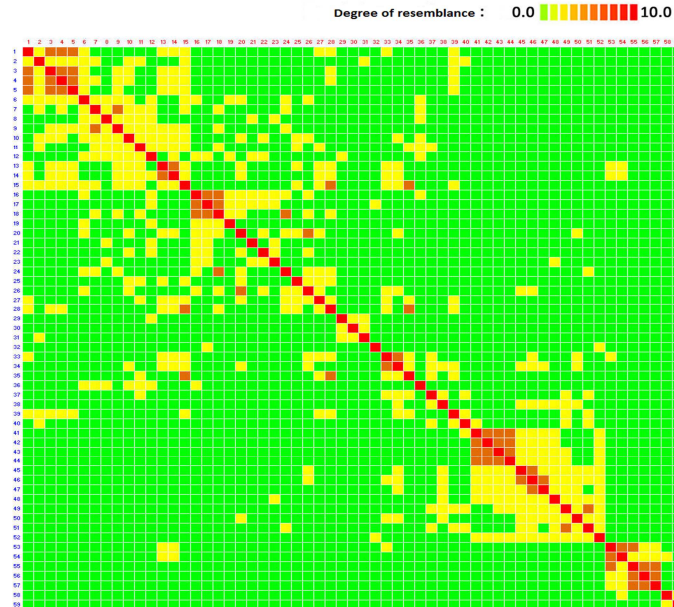


Fig. 3. Similarity matrix corresponding to the hand shapes.

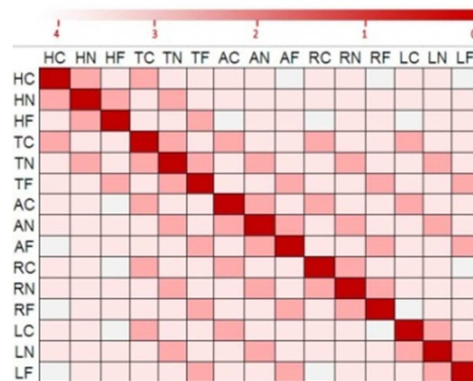


Fig. 4. Similarity matrix corresponding to the hand positions.

During the first phase of our experiments, we conducted preliminary tests in order to define an experimental protocol on the signs. We have implemented a web interface that allows the selection of a hand shape and assigns to it a position in space. Next, we investigated the validity of the results with respect to the input data and the corresponding signs proposed. The first results of this experiment showed that two parameters are insufficient to find a particular sign. We found a success rate of 55 % which is rather low.

The second stage of this research was, hence, to add a third parameter: the movement of the hand. Thus, we defined 10 hand movements that encompass all possible movements in the FSL. Figure 5 shows the distribution of these movements into two categories. The first one corresponds to the movements which can be performed by a single hand, while the second one corresponds to those that require the use of both hands simultaneously or alternately. Note also that some movements can be described through the combination of movements of the first class exercised simultaneously by two hands (Example: iteration with two hands).

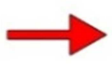
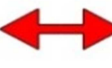

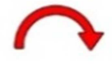

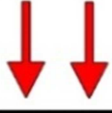
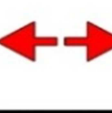
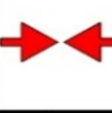


| | | | | | |
|-----------|---|---|---|--|---|
| One hand |  |  |  |  |  |
| | Straight line | Iteration | Wave | Curve | Rotation |
| Two hands |  |  |  |  |  |
| | Double | Separation | Union | Square | Circle |

Fig. 5. The movement of hand during the sign.

We have also taken into account the signs that occur without movement. In total, we used eleven elements to represent all the possible hand movements. The similarity matrix corresponding to this parameter is shown in figure 6.

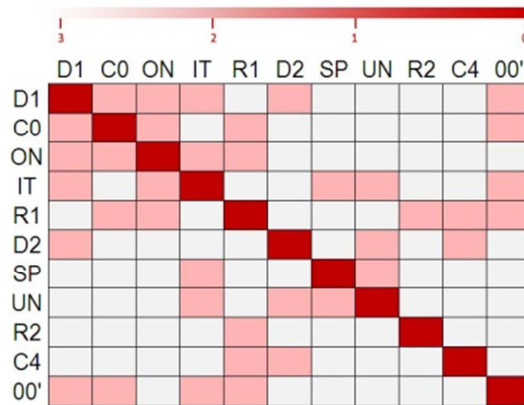


Fig. 6. Similarity matrix corresponding to the hand movements.

In order to better filter our results, we decided to add another element in the coding. This element is related to the number of hands employed. This addition allows us to better discriminate the different signs when the database becomes larger.

Finally, each sign can be represented by the four parameters: [C, P, M, N].

With:

C: Hand shape.

P: Position.

M: Movement.

N: the number of hands used.

To illustrate this encoding, consider the example of the sign "Lundi" ("Monday" in French). This sign, which corresponds to the graphical representation shown in figure 7, is encoded by the following parameters:

[C,P,M,N] = ([08],[RN],[C0],[01]).

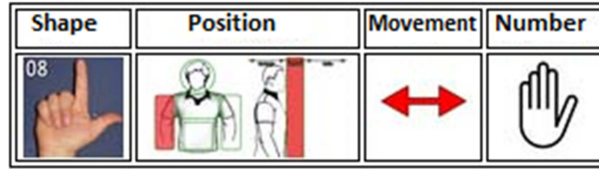


Fig. 7. Similarity matrix corresponding to the hand movements.

3 Classification and results

Using the previous parameters, we preceded to the normalization of the different similarity matrices. We propose to use the following function, which takes into account the global reference symbols and the encodings of the users. The latter is based on the values of the three similarity matrices which correspond to the correct result. To do that, we decided to operate with the matrices initialized and normalized by the following formula:

$$DIS(S_{ref(i)}, S_{user(i)}) = \sum_i \alpha (dis(C_{ref(i)}, C_{user(i)})) + \beta (dis(P_{ref(i)}, P_{user(i)})) + \gamma (dis(M_{ref(i)}, M_{user(i)})) + \delta (dis(N_{ref(i)}, N_{user(i)}))$$

Where :

$DIS(S_{ref(i)}, S_{user(i)})$: Global dissimilarity.

$dis(C_{ref(i)}, C_{user(i)})$: Dissimilarity between hand shapes.

$dis(P_{ref(i)}, P_{user(i)})$: Dissimilarity between positions.

$dis(M_{ref(i)}, M_{user(i)})$: Dissimilarity between movements.

$dis(N_{ref(i)}, N_{user(i)})$: Dissimilarity between hand numbers.

$\alpha, \beta, \gamma, \delta$: Learning coefficients with $\alpha + \beta + \gamma + \delta = 1$.

We introduced this formula with four learning coefficients assigned to the four used parameters. These coefficients are estimated to fit the experiments. Before starting the experiments, we generated the confusion matrix signs using the previous

formula (with $\alpha = \beta = \gamma = \delta$). The measure of dissimilarity between all signs gives a symmetric diagonal matrix zero $M[i,j]$. However, if the matrix $M[i,j]$ is null when $i \neq j$, this indicates a confusion of the two signs i and j in the representation space. This means similar coding of the signs i and j despite their visual difference. Such difference can exist in other settings (facial expressions), which we have not used in this study due to the ergonomic constraints.

After defining our calculation function, we launched our system for experimental tests. Twenty participants contributed to this experiment. These participants have different levels in the practice of FSL. Throughout this step, participants made a total of 6348 observations, which we subsequently filtered in the analysis phase by removing answers too far from reality. Note that the experiments were carried out with static and equal coefficients. The main objective is to define the coefficients values that can fully optimize the results.

Finally, we fixed the more appropriate formula for our system by sorting the classification results. This formula is the following:

$$[(0.41 * C) + (0.27 * P) + (0.23 * M) + (0.09 * N)]$$

The obtained coefficients optimize the maximum number of replies and the success rate for our system. By analyzing these learning coefficients, we observe that the fourth coefficient (number of hands) plays a minimal role. We believe that such parameter can be used at the end to filter the results by redirect the responses to the correct class of signs.

The developed application allows the user to model easily and quickly the sign that he wants to know the meaning of. Once the various parameters are entered, the sign is coded. The results of this coding are presented to the user in the form of a wall of words (Figure 8). The system sorts the signs according to their degree of similarity. The display is based on the results: the font size of the words displayed is related to the level of correspondence to the coding signs.



Fig. 8. Wall of words according to the degree of resemblance.

4 Conclusion

The presented model has allowed us to build a flexible and adaptive architecture to search for a sign. The platform collects user queries by adjusting the learning coefficients and coding signs in the database.

Currently we only consider the observations of the subjects that participated in our experiments. That has allowed us to control the coding and avoid possible interference consultations. However, in the future, we propose to work on the generalization of the system so that, with time, it can self-manage and expand by itself.

The system is available to the public on our website to perform sign searches: <http://www2.univ-paris8.fr/ingenierie-cognition/master-handi/di/>

References

1. C. Cuxac, La Langue des Signes Française : les Voies de l'Iconicité, Faits de Langues Vol. 15-16, Ophrys, 2000.
2. Lefebvre-Albaret F., Dalle P., "Body posture estimation in a sign language video", congress Gesture Workshop, Bielefeld, 2009.
3. A. Braffort, "ARGo: An architecture for sign language recognition and interpretation". In: "Progress in Gestural Interaction", International Gesture Workshop (GW'96), P. Harling and A. Edwards (Eds.), Springer 1997.
4. G. Aznar. Informatisation d'une forme écrite de la langue des Signes Française, Phd thesis, Université de Toulouse, 2008.
5. La langue des Signes 1, 2 et 3. International Visual Theatre Edition, 1998.
6. L. Boutora, Vers un inventaire ordonné des configurations manuelles de la Langue des Signes Française, JEP, Dinard, 2006.
7. F. Bonnal. Sémiogenèse de la langue des signes française : étude critique des signes attestés sur support papier depuis le XVIIIe siècle et nouvelles perspectives de dictionnaires, Thèse de doctorat, Université Toulouse le Mirail, 2005.
8. C. Cuxac, La Langue des Signes Française : les Voies de l'Iconicité, Faits de Langues Vol. 15-16, Ophrys, 2000.

Article 3

Tête parlante codeuse en LPC pour les sourds et les
malentendants

Handicap 2010

Tête parlante codeuse en LPC pour les sourds et les malentendants

M. ZBAKH, H. DAASSI-GNABA, J. LOPEZ KRAHE
Laboratoire Technologies, Handicaps, Interfaces et Multimodalités,
EA 4004 CHART, Université Paris8
2 rue de la Liberté, 93526 Saint Denis Cedex, France
hela.daassi@univ-paris8.fr

Résumé : Notre travail porte sur le développement d'une interface homme machine permettant d'augmenter l'autonomie des sourds et malentendants. Notre principal objectif consiste à intégrer la modalité du Langage Parlé Complété (LPC) dans une tête parlante 3D audiovisuelle, nommée Greta [1].

Cette tête parlante Greta est capable, à partir d'une chaîne phonétique, de générer un signal audio synthétique et les mouvements correspondants des lèvres. Afin de supprimer d'éventuelles ambiguïtés inhérentes à la lecture sur les lèvres, ce système peut être couplé par la vision des mouvements de la main de Greta reproduisant ainsi les gestes du LPC. De ce fait, nous avons développé un module des codes (LPC) puis nous avons implémenté ce module dans la plate forme Greta. Les résultats sont prometteurs et montrent clairement un apport d'information de la parole codée en LPC.

Mots clés : Tête parlante Greta, Langage Parlé Complété (LPC), animation 3D, sourds et malentendants.

I. INTRODUCTION

La lecture labiale est une technique qui permet de comprendre ce que dit une personne en lisant sur ses lèvres. Le LPC (Langage Parlé Complété) hérité du Cued Speech (CS) [2] complète parfaitement la lecture labiale. Cela donne une perception visuelle précise de la parole : la main du locuteur, placée près du visage complète le mouvement des lèvres, permettant ainsi de lever l'ambiguïté existant entre plusieurs phonèmes correspondant au même mouvement des lèvres. En Français cinq positions de la main sont utilisées pour coder les voyelles et huit formes de main sont utilisées pour les consonnes (Fig .1). La combinaison d'une

configuration et d'une position code une syllabe de type Consonne-Voyelle CV [3] [4].

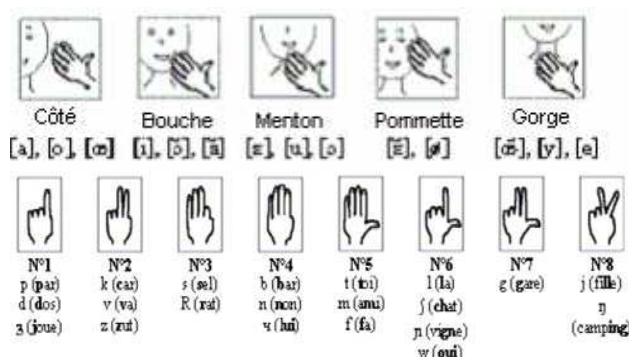


Fig. 1 : Liste des positions et des configurations manuelles du code LPC (issue de [3]).

Il convient de préciser que le LPC est un système de codage associé au langage parlé très différente de la Langue de Signes (LS), langue pratiquée habituellement pour communiquer par les personnes sourdes. La LS dispose d'une grammaire, un vocabulaire, une syntaxe spécifique. Une structure de linguistique particulière très différente de langues parlées, d'où la difficulté d'un traitement automatisé (traduction automatique, par exemple).

Le travail présenté dans cette communication entre dans le cadre de la réalisation d'un système combiné : reconnaissance de la parole, reconnaissance des émotions et tête parlante [5]. Ce système a pour rôle de reconnaître les phonèmes prononcés par le locuteur et son état émotionnel afin d'animer, sur un écran informatique, une tête parlante située devant la personne sourde ou malentendante. Cette tête de synthèse [6] [7] [8], dans notre cas nommée Greta [1] produit des expressions faciales et des mouvements de lèvres. Ce système peut être amélioré par la vision des mouvements de la main de Greta reproduisant ainsi

les gestes du LPC. C'est la raison pour laquelle nous avons développé un module des codes LPC puis nous avons implémenté ce module dans la plateforme Greta.

Nous décrivons l'architecture et la structure de Greta avec une description de l'interface homme machine spécifique qui a été développée pour le codage en LPC. Nous finirons avec une présentation et interprétation des résultats obtenus. Des conclusions et perspectives clôturent notre communication.

II. LA TETE PARLANTE

A. Description de l'architecture de Greta

Greta est un système multimodal capable de communiquer verbalement et non verbalement (Fig .2). Ainsi, il peut générer un comportement visuel et expressif. Ce système est divisé en deux grands modules : le premier module s'intéresse au visage. La forme, la texture et les expressions du visage sont généralement commandées par le bitstream contenant des exemples des ensembles FDP (Facial Definition Parameter) paramètres de définition faciaux, et FAP (Facial Animation Parameter) paramètres de l'animation faciaux qui gèrent les mouvements du visage. Le deuxième module contient tous les fichiers qui interviennent dans le contrôle des gestes du corps dans ce cas là, il s'agit BAP (Body Animation Parameter) [9].



Fig .2 : Greta : visage (les lèvres) et corps (les mains).

Dans cette étude, nous nous sommes concentrés sur l'animation des mouvements des lèvres pour les expressions faciales, et les mouvements de la main pour le comportement gestuel.

B. Le choix de Greta

Les Agents Conversationnels Animés (ACAs) sont particulièrement intéressants dans le contexte d'études expérimentales sur la perception de comportements émotionnels multimodaux, parce qu'on peut contrôler leurs signaux comportementaux et même leurs modalités.

Il est important de pouvoir associer des émotions à l'expression du visage. Ceci est une demande des personnes qui regardent ce type d'avatars. Un visage sans expression devient rapidement ennuyeux et peut annuler tout un travail de recherche de perfection dans le système de codage.

Notre choix pour Greta était basé sur le fait que cette tête parlante, en plus des gestes et des mouvements des lèvres, peut s'exprimer par des émotions et des expressions faciales. Ces dernières sont appelées par des mots clés selon l'état émotionnel du locuteur (« Joy », « Fear », « Anger »...) [9] (Fig .3).

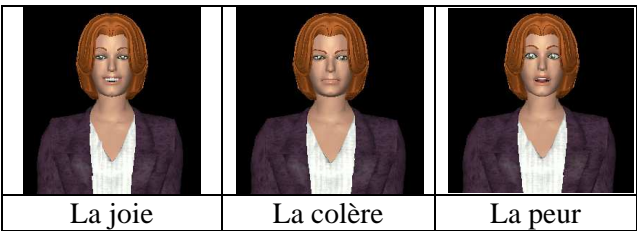


Fig .3 : Expressions faciales de Greta pour différentes émotions.

III. GRETA CODEUSE EN LPC

Nous avons développé une interface pour le codage des gestes du LPC. Cette interface nous a permis de créer les huit configurations du LPC pour coder les consonnes, ainsi les cinq positions pour coder les voyelles [3].

Pour le développement de notre interface, nous avons utilisé la main droite de Greta comme une main dominante (Fig .4.1). Ce choix était basé sur le fait que la majorité des personnes sont droitrières.

Cette interface appelle des fichiers contenant tous les paramètres de configurations. La figure 4.2 montre un exemple de structure d'un fichier qui correspond à la configuration N°5 (Fig .4.1). Ce fichier texte contient trois colonnes qui correspondent aux coordonnées (X, Y et Z) des points. Ces derniers définissent les gestes LPC de la main dans l'espace. Il s'agit de vingt points :

- le **poignet** est présenté par un seul point (pris comme origines);
- Le **pouce** est codé par trois points;
- L'**index**, le **majeur**, l'**annulaire** et l'**articulaire** sont présentés par quatre points;



Fig .4.1 : Points de contrôle de la main droite (conf. N° 5).

| | X | Y | Z |
|-----------|--------|------|-------|
| r_wrist | 0.00 | 0.00 | 0.00 |
| r_thumb1 | -14.00 | 2.00 | -2.00 |
| r_thumb2 | -7.00 | 0.00 | 0.00 |
| r_thumb3 | -16.00 | 0.00 | 0.00 |
| r_index0 | 0.00 | 0.00 | 0.00 |
| r_index1 | 2.00 | 0.00 | -2.00 |
| r_index2 | 0.00 | 0.00 | 0.00 |
| r_index3 | 0.00 | 0.00 | 0.00 |
| r_middle0 | 0.00 | 0.00 | 0.00 |
| r_middle1 | -2.00 | 0.00 | -5.00 |
| r_middle2 | 0.00 | 0.00 | 0.00 |
| r_middle3 | 0.00 | 0.00 | 0.00 |
| r_ring0 | 0.00 | 0.00 | 0.00 |
| r_ring1 | -7.00 | 0.00 | -7.00 |
| r_ring2 | 0.00 | 0.00 | 0.00 |
| r_ring3 | 0.00 | 0.00 | 0.00 |
| r_pinky0 | 0.00 | 0.00 | 0.00 |
| r_pinky1 | -11.00 | 0.00 | -9.00 |
| r_pinky2 | 0.00 | 0.00 | 0.00 |
| r_pinky3 | 0.00 | 0.00 | 0.00 |

Fig .4.2 : Exemple de structure de fichier pour (la conf. N° 5).

Après la création des fichiers de configurations et l'affectation des valeurs de positions dans le code de Greta, nous avons obtenu les résultats illustrés par la figure 5.

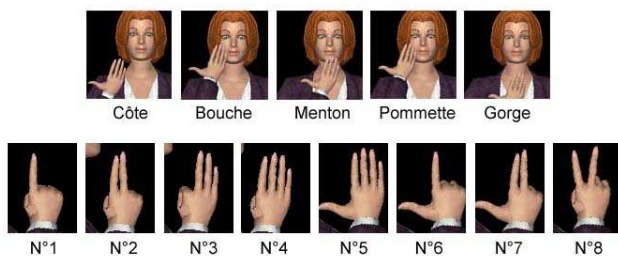


Fig .5 : Liste des positions et des configurations manuelles du code LPC (main droite de Greta).

Comme indiqué précédemment (section 1), dans le LPC, chaque configuration doit être combinée à une position. Cette combinaison code une syllabe de type Consonne-Voyelle CV. Nous avons créé les quarante combinaisons de type Configuration-Position du LPC, qui correspondent aux syllabes (CV) de la langue française. Ceci a pour but de simplifier et accélérer l'affectation de chaque syllabe au geste du LPC correspondant.

Le rôle essentiel de cette interface est l'enregistrement des gestes du LPC, afin de les exécuter au moment de l'animation de Greta (cf. Fig .6).

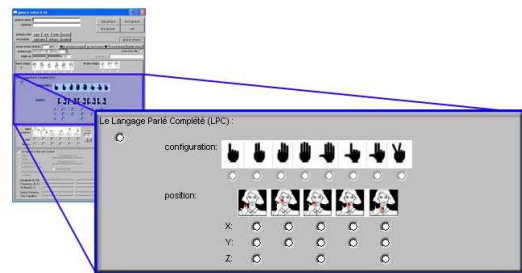


Fig .6 : Nouvelle interface de Greta contenant des gestes du LPC.

IV. INTERPRETATION ET DISCUSSION DES RESULTATS

A. Gestes du LPC

Généralement, Greta utilise un fichier XML [9] pour exécuter les gestes. Dans notre cas, il s'agit des gestes spécifiques (du LPC) qui ont été enregistrés par notre interface (Fig .6), décrite dans la section précédente.

Il y a deux méthodes d'utilisation du fichier XML : la première consiste à citer tous les gestes du LPC dans l'ordre de l'exécution et la deuxième consiste à mentionner le contenu linguistique de la phrase à prononcer.

Dans le test présenté ici, nous avons utilisé la première méthode. Nous avons choisi la phrase "Elle a un piano à pile, et plus de pile d'ailleurs...". Le fichier XML utilisé est composé d'une succession des balises « gesture ». Ces balises contiennent les codes du LPC qui correspondent aux syllabes de la phrase choisie. Soit le code suivant qui présente un extrait de ce fichier XML. Il s'agit d'une balise « gesture » qui contient le nom du geste à exécuter, le moment du démarrage et la durée de l'exécution.

```

...
<gesture id="g2" start="1.3" end="0.50" stroke="0.2">
  <description level="1" type="gretahml">
    <reference>LPC=Conf6Cote</reference>
    <intensity>1.00</intensity>
    <FLD.value>0.00</FLD.value>
    <PWR.value>0.00</PWR.value>
    <REP.value>0.00</REP.value>
    <SPC.value>0.00</SPC.value>
    <TMP.value>0.00</TMP.value>
  </description>
</gesture>
...

```

Cette partie du code correspond au geste « Conf6Cote » : c'est la configuration N°6 associée à la position Côté (Fig .5). Ce geste commence après 1,3 seconde du lancement de l'animation et dure 0,5 seconde. Dans cet exemple, ce geste « Conf6Cote » code la syllabe « la ».

Nous notons que ce test a été effectué sans synthèse vocale : sans animation des mouvements des lèvres de Geta.

Les résultats, les gestes du LPC générés par Greta pour coder toutes les syllabes de la phrase choisie, sont présentés sur la figure 7. Afin d’avoir une comparaison objective, la même phrase a été prononcée par une animatrice [10]. Cette dernière parle et effectue simultanément les gestes du LPC.

Le but de cette comparaison est de vérifier la clarification des gestes du LPC générés par Greta et la compatibilité de ces gestes avec ceux de l’animatrice.

| Animatrice | Greta |
|---|---|
|  |  |
| Elle a un piano à pile, et plus de pile d'ailleurs... | |
|  |  |
| Elle a un piano à pile, et plus de pile d'ailleurs... | |
|  |  |
| Elle a un piano à pile, et plus de pile d'ailleurs... | |
|  |  |
| Elle a un piano à pile, et plus de pile d'ailleurs... | |
|  |  |
| Elle a un piano à pile, et plus de pile d'ailleurs... | |

| | |
|---|--|
|  |  |
| Elle a un piano à pile, et plus de pile d'ailleurs... | |
|  |  |
| Elle a un piano à pile, et plus de pile d'ailleurs... | |
|  |  |
| Elle a un piano à pile, et plus de pile d'ailleurs... | |
|  |  |
| Elle a un piano à pile, et plus de pile d'ailleurs... | |

Fig .7 : Comparaison entre les gestes du Greta et ceux de l’animatrice.

D’après la figure 7, nous avons comparé deux paramètres : le premier paramètre est le degré de ressemblance entre les configurations de Greta et celles de l’animatrice. Le deuxième paramètre est le temps de l’exécution de ces configurations. Les résultats montrent qu’avec la méthode proposée dans cette section, Greta ne présente pas une différence des gestes du LPC par rapport à ceux effectués par l’animatrice, soit au niveau de ressemblance, soit au niveau de temps (Fig .7).

Le fichier XML organise la création des gestes d’une manière efficace. Cette organisation est due à la temporisation de chaque configuration, ainsi sa durée de l’exécution. Cette méthode demande toujours une intervention de l’administrateur. La création du fichier XML et l’affectation des valeurs se font manuellement, ce qui complique la gestion de ces fichiers et limite le champ de l’exploitation.

B. Gestes du LPC et mouvements des lèvres

Dans ce test, nous avons essayé de résoudre la difficulté de la gestion des fichiers XML par l'utilisation de la deuxième méthode, mentionnée dans la section 4.1.

En effet le contenu est remplacé par la phrase textuelle au lieu des balises des gestes. La phrase prononcée par Greta est la même : "*Elle a un piano à pile, et plus de pile d'ailleurs...*". Ce test a été accompagné par une synthèse vocale française nommée WinEuler [11].

Durant ce test, Greta est devenue plus « autonome ». L'animation des gestes du LPC est effectuée par une traduction phonème-geste de la phrase restituée.

En effet, la synthèse vocale crée un fichier phonétique qui contient tous les phonèmes de la phrase textuelle. Cette synthèse vocale segmente la phrase textuelle en phonèmes avec un symbole spécifique pour chaque phonème. Après la création de ce fichier, Greta extrait la liste des phonèmes à l'aide de la fonction « GestureLPC » que nous avons développée. Le rôle de cette fonction est de parcourir le contenu du fichier pour attribuer chaque phonème au geste correspondant. Chaque phonème correspond à une configuration ou à une position du LPC.

La fonction « GestureLPC » crée un vecteur de configurations qui correspond exactement à la liste des phonèmes générée par la synthèse vocale. Au moment de l'animation, l'application Greta lance la synthèse vocale et l'exécution des gestes du LPC. Cette démonstration est effectuée en 3 étapes : la première concerne la création du fichier phonétique par la synthèse vocale de Greta, la deuxième consiste à l'élaboration de la liste des gestes du LPC à partir du fichier phonétique et la dernière correspond à l'exécution de la synthèse vocale et des gestes du LPC en même temps. Les résultats montrent une génération de la synthèse vocale accompagnée par les expressions gestuelles (codage du LPC). Dans cette méthode la gestion de l'animation s'effectue automatiquement grâce à la fonction « GestureLPC ».

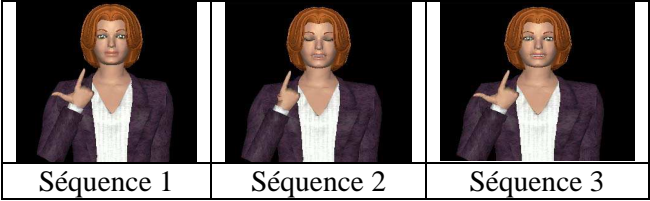


Fig .8 : Démonstration des gestes du LPC avec les mouvements des lèvres.

Durant ce test, Greta est devenue tête parlante codeuse en LPC (Fig .8). Nous avons constaté que les résultats obtenus complètent ceux de la section 4.1. Il y a les mouvements des lèvres accompagnés par le codage du LPC. La création d'une liste des configurations et des positions à partir du fichier phonétique nous a conduit à travailler sur la synchronisation entre les mouvements des lèvres et les gestes du LPC. Vu que la fonction « GestureLPC » ne fusionne pas les configurations avec leurs positions. Par la suite, Greta anime séparément les configurations et les positions. C'est pour cela que notre future tâche est de composer des clés de genre configuration/position pour coder les syllabes de la langue française.

C. Gestes du LPC, mouvements des lèvres et expressions faciales

Des éléments complémentaires ont été ajoutés dans l'animation afin d'améliorer la présentation de la tête parlante. Nous parlons des expressions faciales.

Dans ce test, la tête parlante anime les gestes du LPC combinés avec les mouvements des lèvres et les émotions. La phrase que nous avons considérée, dans ce test, est "*La partie belotte dura toute la journée.*". Nous présentons sur la figure 9 les expressions faciales de Greta (codeuse en LPC) qui correspondent aux émotions : joie, colère et peur.



Fig .9 : Expressions faciales de Greta codeuse en LPC pour différentes émotions.

A travers les résultats illustrés dans la figure 9, nous avons remarqué que la combinaison des gestes

du LPC avec les mouvements des lèvres et les expressions faciales donnent une vue plus animée de la tête parlante Greta.

Cette interface animée Greta codeuse en LPC complète maintenant le système combiné [5] : reconnaissance de la parole, reconnaissance des émotions à partir d'un signal audio pour animer une tête parlante dédiée aux personnes sourdes et malentendantes. Puisque les utilisateurs de ce système exigent non seulement le contenu linguistique de la parole mais aussi l'état émotionnel du locuteur. Ceci s'explique par le fait que les émotions jouent un rôle fondamental pour l'identification du contexte de la parole.

V. CONCLUSION ET PERSPECTIVES

Dans ce papier nous avons présenté trois versions de la tête parlante Greta : une première version respecte le codage du LPC concernant les configurations et les positions, ainsi, la rapidité de l'animation. Cette version n'applique pas la synthèse vocale : La tête parlante effectue les gestes du LPC mais sans aucun mouvements des lèvres. La deuxième version est plus utile. La tête parlante peut prononcer la synthèse de la parole avec la possibilité d'ajouter les gestes du LPC. La troisième version fusionne trois éléments : les gestes du LPC, la parole et les expressions faciales. Dans cette dernière version nous avons remarqué le rôle important des émotions par rapport aux deux autres versions.

Nos perspectives d'avenir seront concentrées sur notre outil du codage LPC et sur la synchronisation de la synthèse vocale. C'est pour cela, nous sommes en train d'intégrer un algorithme de programmation dynamique, proche de **Levenshtein Distance** [12] [13], pour assurer la mise en correspondance du codage syllabique adapté à celui du LPC.

Nous avons remarqué que l'articulation des phonèmes devrait être plus claire pour la lecture labiale. Un futur travail sera concentré sur l'optimisation des mouvements des lèvres. Afin de rendre cette articulation plus réelle, nous proposons alors de travailler directement sur les syllabes. Dans ce type de développement, le codage du LPC permet également de mieux compléter la lecture labiale.

VI. REMERCIEMENTS

Nous tenons à remercier tous les membres du laboratoire LTCI de Télécom ParisTech et

particulièrement Mme Catherine Pelachaud et M Radek Niewiadomski.

VII. BIBLIOGRAPHIE

- [1] C. Pelachaud, V. Carofiglio, B. De Carolis, F. De Rosis, and I. Poggi, "Embodied contextual agent in information delivering application", *First International Joint Conference on Autonomous Agent and Multiagent System*, pp. 758 -765, 2002.
- [2] R.O. Cornett, "Cued Speech", *American Annals of the Deaf*, Vol. 112, pp. 3-13, 1967.
- [3] V. Attina, D. Beautemps, M. A. Cathiard, and M. Odisio, "A pilot study of temporal organization in cued speech production of French syllables: rules for Cued Speech synthesizer", *Speech Communication*, Vol. 44, pp. 197-214, 2004.
- [4] N. Aboutabit, D. Beautemps, and L. Besacier, "Hand and Lips desynchronization analysis in French Cued Speech: Automatic segmentation of Hand flow", *ICASSP*, 2006.
- [5] H. Daassi-Gnaba, and J. Lopez Krahe, "Universal combined system: speech recognition, emotion recognition and talking head for deaf and hard of hearing people", *AAATE*, pp. 503-508, 2009.
- [6] I. Williams, "ALLAO, un système d'apprentissage et d'évaluation de la lecture labiale à partir d'un vidéodisque interactif", *Bulletin d'Audiophonologie*, Vol. 3, n° 5, pp. 89-614, 1987.
- [7] L. Revèret, G. Bailly, and P. Badin, "MOTHER: a new generation of talking heads providing a flexible articulatory control for video-realistic speech animation", *ICSLP*, pp. 755-758, 2000.
- [8] G. Gibert, "Conception et évaluation d'un système de synthèse 3D de Langue française Parlée Complétée (LPC) à partir du texte", *thèse de doctorat de l'Institut National Polytechnique*, Grenoble, France, 2006.
- [9] S. Pasquariello , and C. Pelachaud, "Greta A simple facial Animation Engine", *6th Online World Conference on Soft Computing in Industrial Applications, Session on Soft Computing for Intelligent 3D Agents*, 2001.
- [10] Association nationale pour la promotion et le développement de la Langue française Parlée Complétée, [consulté le 25 septembre 2009]. Disponible sur : <http://www.alpc.asso.fr/code01-c.htm>
- [11] The EULER Project, [consulté le 25 septembre 2009]. Disponible sur : <http://tcts.fpms.ac.be/synthesis/euler/>
- [12] V. I. levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions and Reversals", *Soviet Physics Doklady*, 10, February 1966.
- [13] R. Lowrance, and R. A. Wagner. "An extension of the string-to-string correction problem". *Journal of the ACM*, pp. 177-183, April 1975.