

Fais-moi signe !

—

Traducteur de langue visuo-gestuelle

# Sommaire

- Projet idéal
  - État de l'art
  - Personas
  - Arbre de tâches
  - Modalités
  - Diagrammes CARE et de combinaison
- Projet développé
  - MédiaPipe et modèle
  - Evaluations
  - Retours
- Améliorations
- Conclusion
- Démonstration

# Idée

## Traduction de la langue des signes (LS):

- Le logiciel reconnaît les signes et expressions du visage de la personne en face.
- Une synthèse vocale traduit en direct ce qui est exprimé, complété par du sous-titrage.

## Traduction en LS :

- Le logiciel comprend une reconnaissance vocale, et un avatar interprète.
- Des sous-titres sont aussi affichés.

## Choix de toutes les langues :

- Signées : LSF, ASL, LGP...
- Vocales et textes : Français, Anglais, Portuguais...

# Supports

Ce logiciel peut être intégré n'importe où :

- Lunettes AR : en direct (face à face ou visionnage de vidéo), l'utilisateur voit les sous-titres (et l'avatar) et entendrait la synthèse grâce à l'ostéophonie.
- Application téléphone : en filmant, mais aussi en fournissant une vidéo.
- Ordinateur : en fournissant une vidéo, en enregistrant l'écran ou en utilisant la webcam.

# Besoins

- Accessibilité
- Apprentissage des langues des signes
- Communication entre personnes ne pouvant le faire mutuellement en LS
- Échanges entre malvoyants et malentendants

# État de l'art

Machines pouvant traduire des LS développées depuis les années 70

- Plugin Hand Talk : traduction automatique du contenu en ligne en ASL
- SignAll : écriture de LS/audio vers texte (en anglais), en temps réel
- KinTrans : se rapproche de notre projet (LS vers audio + texte)

# Persona



**Nom :** Noé

**Âge :** 23 ans

**Situation familiale :** En couple

**Profession :** Étudiant

**Habite :** Coloc à Grenoble

## ***Personnalité :***

Noé est une personne active, extravertie et calme. Il a un esprit analytique et apprécie le sport pour se changer les idées. Il est très amoureux de sa petite-amie.

## ***Objectifs :***

- Comprendre les membres de la famille malentendante de sa copine entendante
- Communiquer avec eux avec facilité.

## **Entendant**

## ***Habitudes et expertises :***

Utilise régulièrement les réseaux sociaux et suit les actualités sportives depuis son téléphone ou ordinateur. En dehors du français, il ne sait parler qu'anglais.

## ***Frustrations, craintes :***

- Impossibilité de se faire comprendre par sa belle famille
- A peur de ne pas se faire accepter.
- N'a pas le temps d'apprendre la LSF en suivant des cours.

# Persona



**Nom :** Marie

**Âge :** 31 ans

**Situation familiale :** Mariée,  
avec un enfant

**Profession :** Aide-soignante

**Habite :** Maison en banlieue de Grenoble

## ***Personnalité :***

Marie est très patiente. Elle est très impliquée dans sa vie de famille. Elle aime aller au cinéma, ses films préférés sont les Harry Potter. Elle est sourde appareillée.

## ***Objectifs :***

- Apprendre la LSF à son fils qui lui entend bien.
- Mieux communiquer avec son entourage.

## **Sourde**

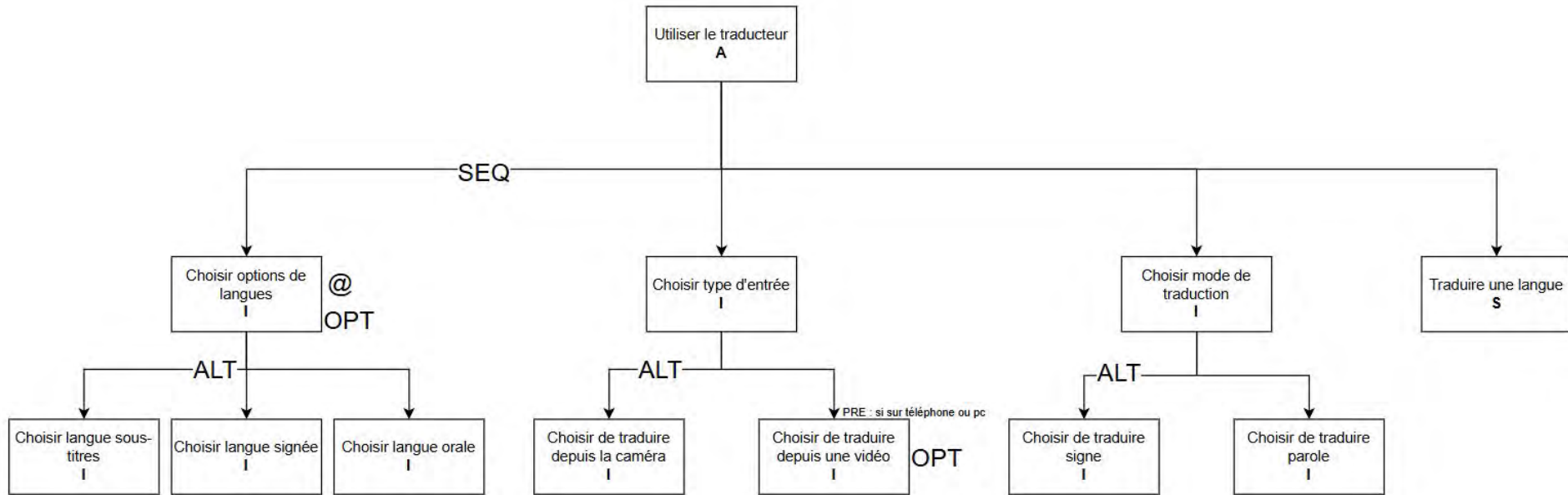
## ***Habitudes et expertises :***

A l'habitude de côtoyer des personnes handicapées. Elle aime passer du temps à scroller sur les réseaux avant de se coucher.

## ***Frustrations, craintes :***

- A parfois du mal à se faire comprendre par ses proches lorsqu'elle ne porte pas ses appareils.
- A peur de devenir dépendante de l'outil pour pouvoir comprendre les autres.

# Modèle de tâches



# Modalités d'entrée

- M1.1 : < Microphone, langage naturel > (parole) RA/Téléphone/Ordinateur
  - M1.2 : < Lecteur vidéo, langage naturel > (parole) Téléphone/Ordinateur
  - M2.1 : < Caméra, gestes > (signes) RA/Téléphone/Ordinateur
  - M2.2 : < Lecteur vidéo, gestes > (signes) Téléphone/Ordinateur
- 
- M3.1 : < Interface tactile, langage de commande > RA/Téléphone
  - M3.2 : < Souris, langage de commande > Ordinateur

Justification : traduction entre deux types de langues et multiples supports + choix d'options

# Modalités d'entrée : Parole prise de vue réelle

- Sens : ouïe
- Modalité : passive
- Publique
- Spatial : environnement
- Temporalité : persistante
- Dimension : 1D (signal audio)
- Dynamique
- Linguistique

*Microphone*

Modalité  
1.1

*Langage naturel*

Justification : but du traducteur -> pouvoir traduire la **parole** vers gestes + texte

# Modalités d'entrée : Parole **fourniture** vidéo

- Sens : ouïe
- Modalité : passive
- Privée
- Spatial : écran
- Temporalité : persistante
- Dimension : 1D (signal audio)
- Dynamique
- Linguistique

*Lecteur vidéo*

Modalité  
1.2

*Langage naturel*

Justification : but du traducteur -> pouvoir traduire la **parole** vers gestes + texte

# Modalités d'entrée : Gestes prise de vue réelle

- Sens : vue
- Modalité : passive
- Publique
- Spatial : champ caméra
- Temporalité : persistante
- Dimension : 3D
- Dynamique
- Linguistique

*Caméra*

Modalité  
2.1

*Gestes*

Justification : but du traducteur -> pouvoir traduire les **gestes** vers audio + texte

# Modalités d'entrée : Gestes fourniture vidéo

- Sens : vue
- Modalité : passive
- Privée
- Spatial : écran
- Temporalité : persistante
- Dimension : 2D
- Dynamique
- Linguistique

*Lecteur vidéo*

Modalité  
2.2

*Gestes*

Justification : but du traducteur -> pouvoir traduire les **gestes** vers audio + texte

# Modalités d'entrée : Manipulation directe **RA**/téléphone

- Sens : toucher
- Modalité : active
- Privée
- Spatial : bouton
- Temporalité : persistante
- Dimension : 2D
- Non-analogue
- Arbitraire

*Interface tactile*

Modalité  
3.1

*Langage de  
commande*

Justification : pouvoir faire les choix d'options et commandes non gestuelles/vocales, multiples supports

# Modalités d'entrée : Manipulation directe **Ordinateur**

- Sens : toucher
- Modalité : active
- Privée
- Spatial : écran
- Temporalité : persistante
- Dimension : 2D
- Non-analogue
- Arbitraire

*Souris*

Modalité  
3.2

*Langage de  
commande*

Justification : pouvoir faire les choix d'options et commandes non gestuelles/vocales, multiples supports

# Modalités de sortie

- Lunettes RA : M4 :  $\langle \text{Casque ostéophone, langage naturel} \rangle$  (audio)  
M5 :  $\langle \text{RA, langage naturel} \rangle$  (texte)  
M6 :  $\langle \text{RA, graphismes 3D} \rangle$  (signes)
- Téléphone/ordinateur :  
M7 :  $\langle \text{Haut-parleur, langage naturel} \rangle$  (audio)  
M8 :  $\langle \text{Écran, langage naturel} \rangle$  (texte)  
M9 :  $\langle \text{Écran, graphismes 3D} \rangle$  (signes)

Justification : double modalité audio+texte ou signes+texte, afin d'améliorer l'accessibilité (réponses aux besoins) et multiples supports

# Modalités de sortie : Audio RA

- Sens : ouïe
- Modalité : passive
- Privée
- Spatial : champ auditif
- Temporalité : persistante
- Dimension : 1D
- Dynamique
- Linguistique

*Casque ostéophone*

Modalité 4

*Langage naturel*

Justification : but du traducteur -> pouvoir traduire les gestes vers **audio** + texte

# Modalités de sortie : Texte **RA**

- Sens : vue
- Modalité : passive
- Privée
- Spatial : champ de vision
- Temporalité : persistante
- Dimension : 2D
- Dynamique
- Linguistique

*Réalité augmentée*

Modalité 5

*Langage naturel*

Justification : but du traducteur -> pouvoir traduire les gestes vers audio + **texte**

# Modalités de sortie : Gestes RA

- Sens : vue
- Modalité : passive
- Privée
- Spatial : champ de vision
- Temporalité : persistante
- Dimension : 3D
- Dynamique
- Linguistique

*Réalité augmentée*

Modalité 6

*Graphismes 3D*

Justification : but du traducteur -> pouvoir traduire la parole vers **gestes** + texte

# Modalités de sortie : Audio Téléphone/Ordinateur

- Sens : ouïe
- Modalité : passive
- Privée/publique
- Spatial : champ auditif
- Temporalité : persistante
- Dimension : 1D
- Dynamique
- Linguistique

*Haut-parleur*

Modalité 7

*Langage naturel*

Justification : but du traducteur -> pouvoir traduire les gestes vers **audio** + texte

# Modalités de sortie : Texte Téléphone/Ordinateur

- Sens : vue
- Modalité : passive
- Privée
- Spatial : écran
- Temporalité : persistante
- Dimension : 2D
- Dynamique
- Linguistique

*Écran*

Modalité 8

*Langage naturel*

Justification : but du traducteur -> pouvoir traduire les gestes vers audio + **texte**

# Modalités de sortie : Gestes Téléphone/Ordinateur

- Sens : vue
- Modalité : passive
- Privée
- Spatial : écran
- Temporalité : persistante
- Dimension : 3D
- Dynamique
- Linguistique

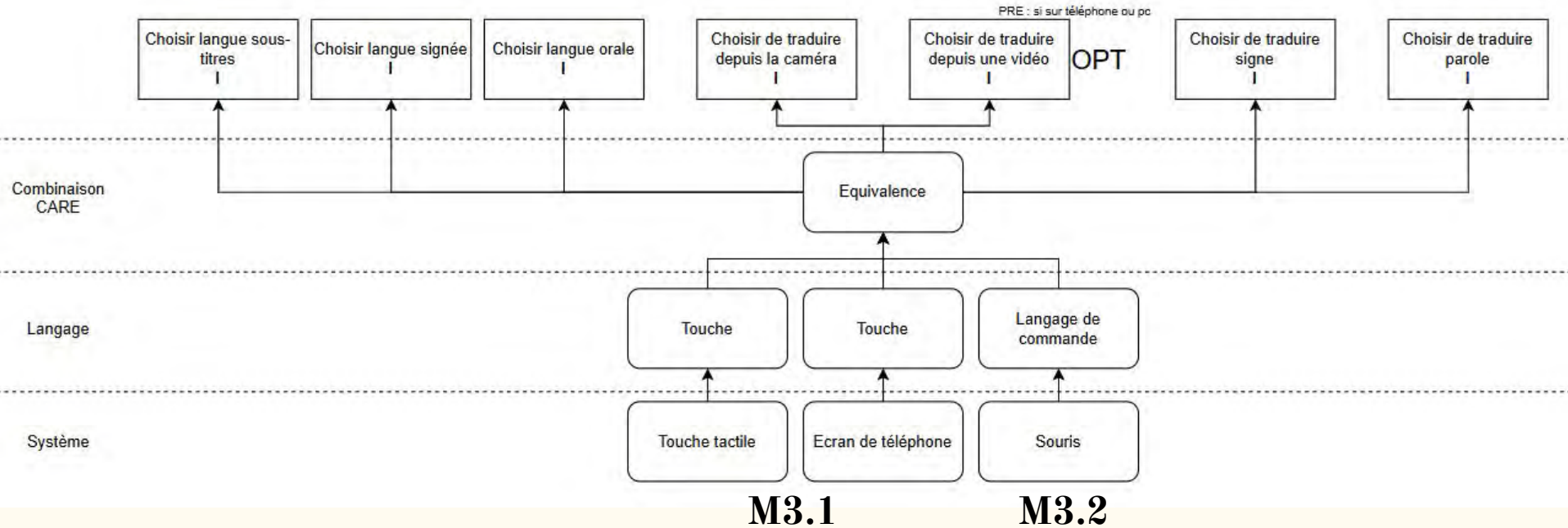
*Écran*

Modalité 9

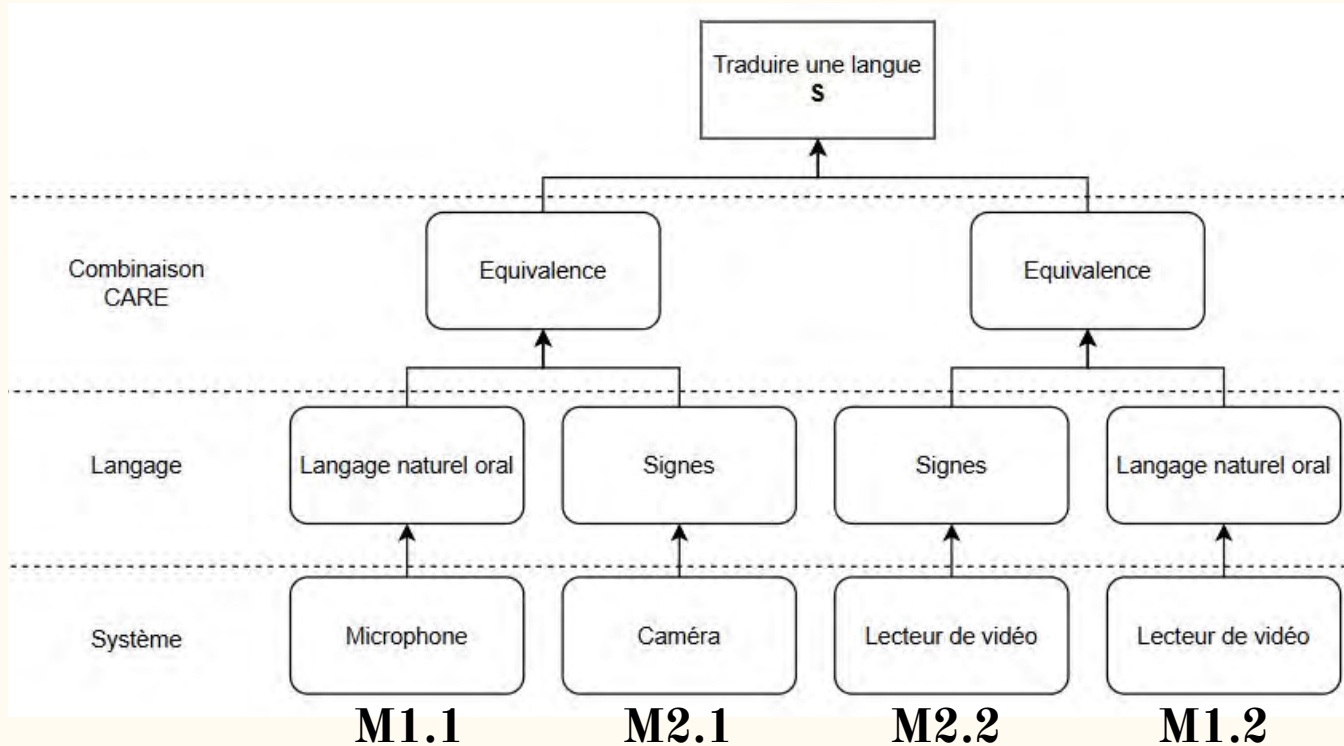
*Graphismes 3D*

Justification : but du traducteur -> pouvoir traduire la parole vers **gestes** + texte

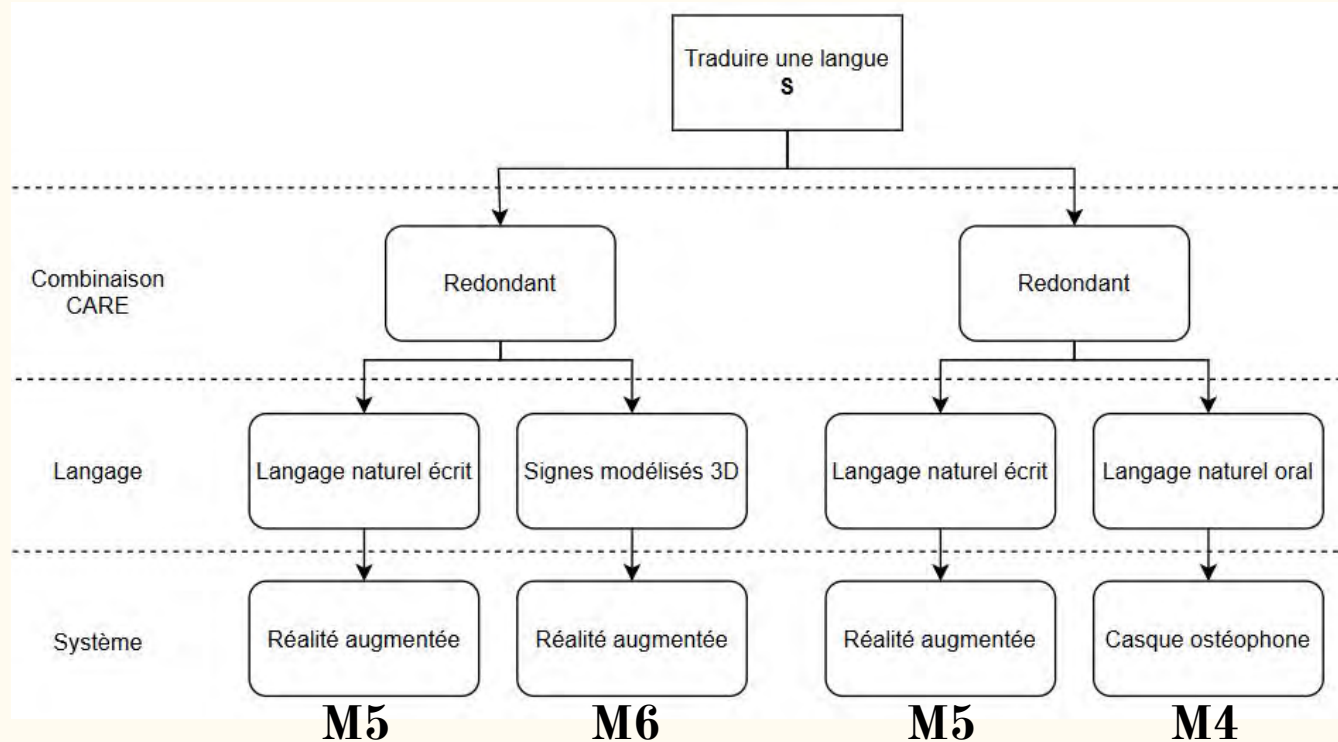
# Diagramme CARE : Entrées



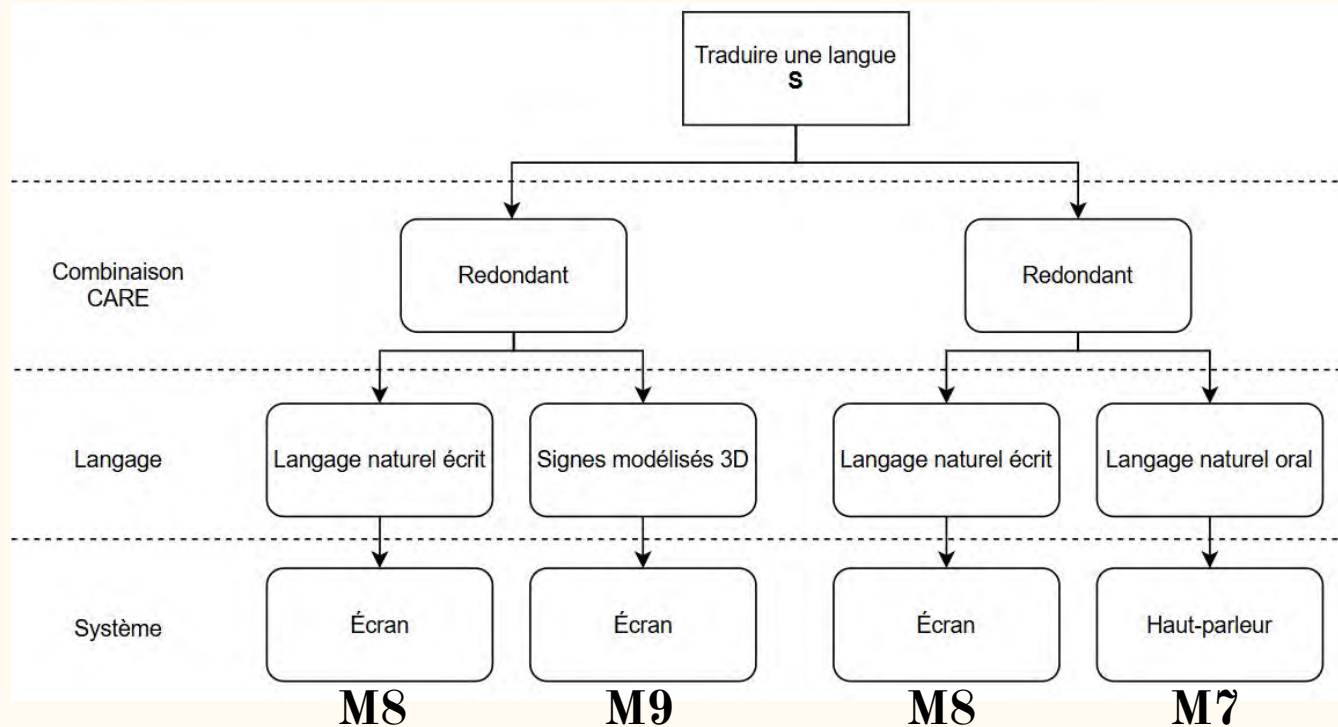
# Diagramme CARE : Entrées



# Diagramme CARE : Sorties RA



# Diagramme CARE : Sorties téléphone/ordinateur



# Diagramme de combinaison

## Combination schemas

Combination aspects

|          |  |  |  |  |  |
|----------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Temporal | Anachronism                                                                       | Sequence                                                                          | Concomitance                                                                        | Coincidence                                                                         | Parallelism                                                                         |
| Spatial  | Separation                                                                        | Adjacency                                                                         | Intersection                                                                        | Overlaid                                                                            | Collocation                                                                         |
| Semantic | Concurrency                                                                       | Complementarity                                                                   | Complementarity & Redundancy                                                        | Partial Redundancy                                                                  | Total Redundancy                                                                    |

$M5 + M4, M5 + M6$

$M8 + M7, M8 + M9$

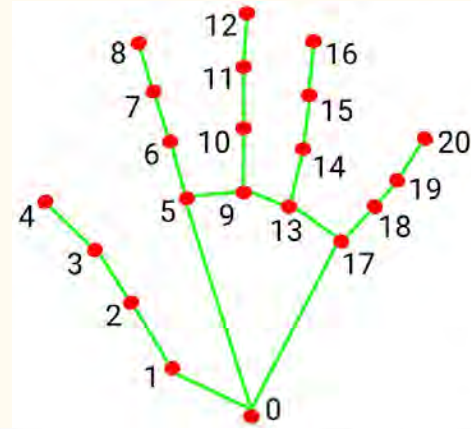
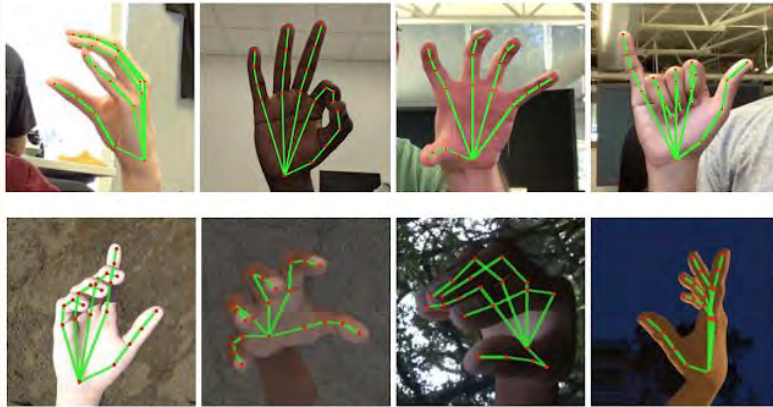
# Projet développé

-> Focus sur la traduction de signes simples :

- Utilisation de la solution **MediaPipe** de Google pour la reconnaissance des signes.
- Entraînement du modèle
- Ajout de sous-titres et voix Text-To-Speech
- En **python**



# MediaPipe



- |                       |                       |
|-----------------------|-----------------------|
| 0. WRIST              | 11. MIDDLE_FINGER_DIP |
| 1. THUMB_CMC          | 12. MIDDLE_FINGER_TIP |
| 2. THUMB_MCP          | 13. RING_FINGER_MCP   |
| 3. THUMB_IP           | 14. RING_FINGER_PIP   |
| 4. THUMB_TIP          | 15. RING_FINGER_DIP   |
| 5. INDEX_FINGER_MCP   | 16. RING_FINGER_TIP   |
| 6. INDEX_FINGER_PIP   | 17. PINKY_MCP         |
| 7. INDEX_FINGER_DIP   | 18. PINKY_PIP         |
| 8. INDEX_FINGER_TIP   | 19. PINKY_DIP         |
| 9. MIDDLE_FINGER_MCP  | 20. PINKY_TIP         |
| 10. MIDDLE_FINGER_PIP |                       |

# Modèle

-> Utilisation de créateur de modèles MediaPipe :

- Bibliothèque alphabet ASL sur Kaggle
- Centaines de photos par lettre
- Des photos ne signant aucune lettre pour montrer le “None”
- Différentes options d'entraînement avec des hyperparamètres (taille de batchs, nombre d'itérations, etc)
- Fichier TensorFlow Lite

# Text-To-Speech

-> Utilisation de gTTS et pyttsx3 :

- gTTS et ses contraintes (WiFi...)
- Linux, pyttsx et longueur
- Text-To-Speech en français grâce à pyttsx et Windows
- Signes parasites => Ajout légère latence

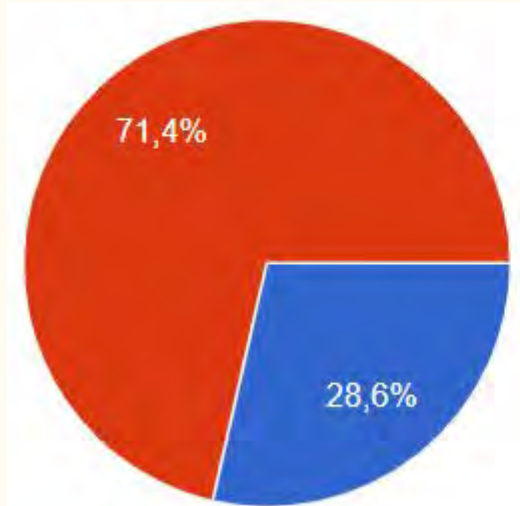
# Notre abécédaire



# Protocole de test

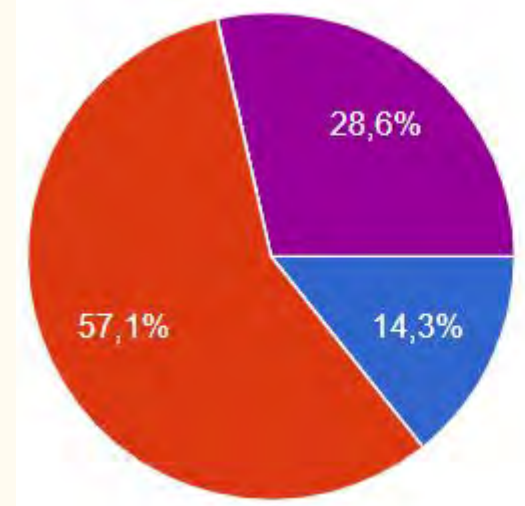
- Explication du projet
- Personne face à l'écran, caméra tournée vers nous de l'autre côté
- Petite leçon sur l'alphabet en ASL :
  - Lui signer des lettres (pas forcément dans l'ordre)
  - Demander quelle lettre on a effectué
- Lui permettre de tourner la caméra pour tester (abécédaire fourni)
- Durant ces temps observation si réactions particulières

# Évaluation : démographie



*Genre*

- Un homme
- Une femme
- Autre / Je ne souhaite pas le préciser

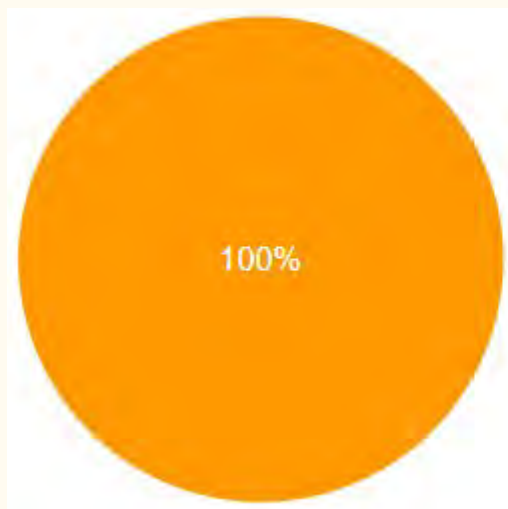


*Âge*

- Moins de 18 ans
- Entre 18 et 25 ans
- Entre 26 et 35 ans
- Entre 36 et 50 ans
- Plus de 50 ans

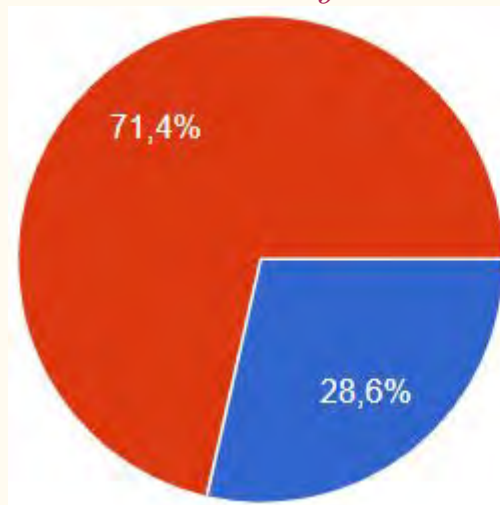
# Évaluation : monde de la surdité

*Malentendant.e ?*



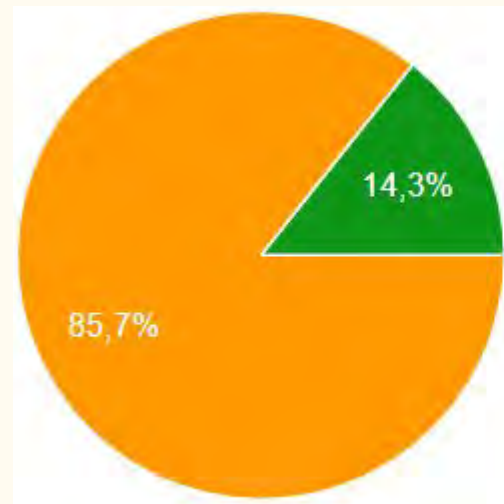
- Oui, et je suis appareillé.e
- Oui, mais je ne suis pas appareillé.e
- Non

*Personne malentendante dans l'entourage*



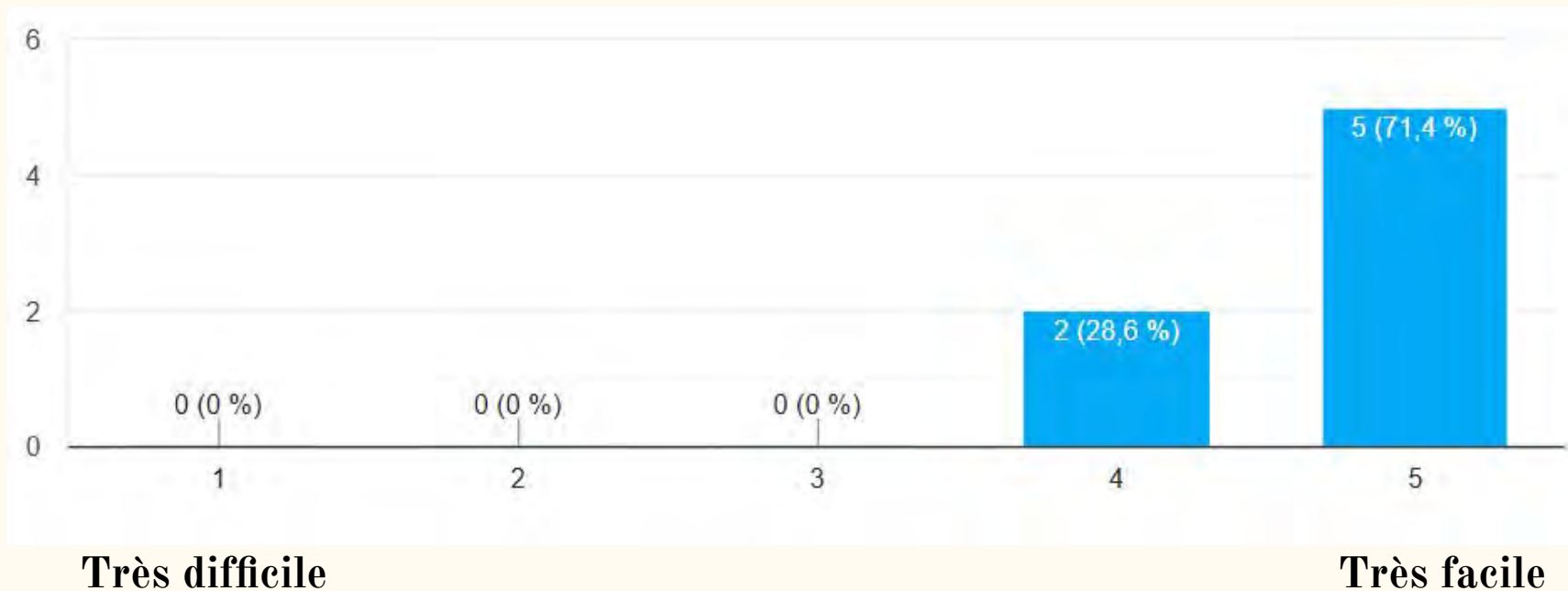
- Oui
- Non

*Pratique d'une LS*



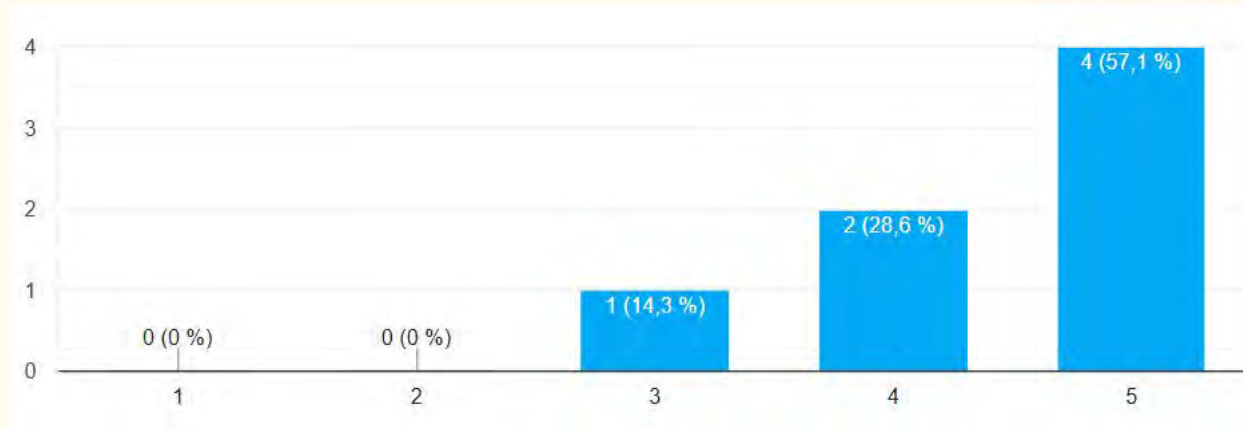
- Oui, régulièrement
- J'ai pu apprendre mais ne l'utilise plus
- Non, mais ça m'intéresse
- Non, et ça ne m'intéresse pas

# Évaluation : reconnaissance des signes



# Évaluation : utilisation future

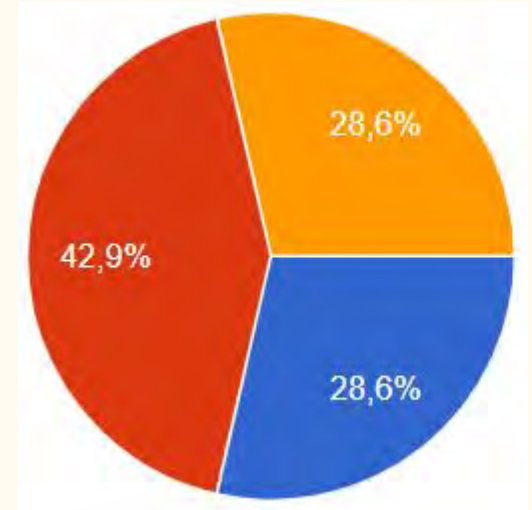
*Utilité de l'application complètement développée*



Pas du tout utile

Très utile

*L'utiliserez-vous ?*



- Oui, très probablement
- Oui, je pense
- Non, je ne pense pas
- Non, pas du tout

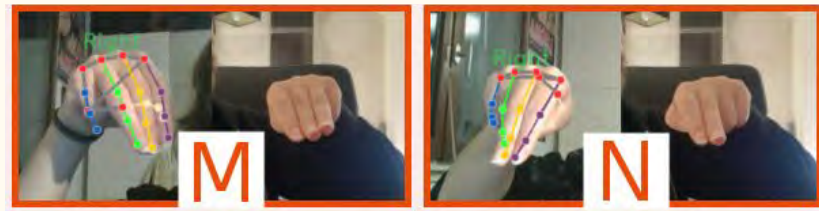
# Observations

- Disent parfois des signes “parasites”
- Ont aimé pouvoir tester par eux-même
- Changement de main pendant les tests
- Tendance à signer dans l’ordre, puis son prénom



# Retours

- **Très positifs** : “bonne appli”, super pour les malentendant.es, intéressant, reconnaissance main droite/gauche appréciée, utile pour apprendre...
- **Utile** lorsque complète (reconnaissance totale et traduction inverse, plusieurs supports)
- **Point faible** :
  - M et N mal détectés (problème connu)
  - Reconnaissance des mots



# Améliorations futur proche

- Ajout reconnaissance du mouvement : permet de signer J et Z et d'autres
- Plus solide reconnaissance (ex : M et N)
- Ajout reconnaissance des expressions faciales
- Enrichissement du modèle
- Ajout de différents modèles entraînés sur différentes langues pour avoir le choix

# Conclusion

- Première base ‘simple’ à améliorer
- Application plus complète pourra être utile
- Retours utilisateurs importants
- Projet très intéressant à imaginer

# Démonstration