A Survey on Mouth Modeling and Analysis for Sign Language Recognition

Epameinondas Antonakos*
Anastasios Roussos*
Stefanos Zafeiriou*

*The authors contributed equally and have joint first authorship. The names appear in alphabetical order.
Outline

- Is mouth linguistically significant for Sign Language?
- Future work, challenges and potentials.
Outline

• Is mouth linguistically significant for Sign Language?

• Review of existing methods of mouth modeling for Automatic Sign Language Recognition.

• Future work, challenges and potentials.
Sign Language (SL)

- About 5% of the worldwide population suffers from hearing loss to some degree.
- 1% of the worldwide population use SLs as their native languages (~70 million deaf people).
- SLs are also used from people who cannot physically speak (mutism).
- There is not a unique international SL. Each country has its own, so there are hundreds of different SLs.
Automatic SL Recognition

- Deaf people encounter many difficulties in the everyday life (education, work, use of the internet, etc.):
  - Limited reading/writing skills in the spoken language (for them it is a foreign language with fundamentally different grammatical structure).
  - The vast majority of the rest of the population is unable to use SL.

- *Automatic SL Recognition (ASLR)* can greatly support the Deaf community.

- However, it is still far from being mature technology, especially compared to text-based interaction or speech recognition.
SL Structure

- SLs are as rich and grammatically complex as spoken languages.
- Manual articulators
  - Phonemes (basic semantic SL components)
  - Hands shape, posture, location and motion
- Non-manual articulators
  - Prosody, lexical distinction, grammatical structure, adjectival/adverbial content
  - Head and body pose, facial expressions (through eyes, eyebrows, cheeks, lips), mouth movements
- Mouth is one of the most involved parts of the face in non-manuals.
Mouth Actions

- Mouth lexical articulators are separated in:

  1) Mouth gestures:
     - Non-verbal components
     - Shape deformation, tongue movement, teeth visibility

  2) Mouthings:
     - Silent articulators that correspond to a pronounced word or part of it.
     - Visual syllables (in most SLs only the first syllable of a word is articulated).
Mouth Actions

- Some argue that mouth actions (especially mouthings) are not linguistically significant.

- Recent research has shown that they contribute significantly to the semantic analysis of SLs.
Mouth Actions

- The frequency of mouth actions is different for each SL.
- It depends on both the context and the grammatical category of the manual sign they occur with.
- Most mouth actions have a *prosodic* interpretation while others have *lexical* meaning.
- In some cases, the mouth articulates physical events, emotions or sensations (types of sounds, noise, disturbances, heaviness, types of textures etc.).
Mouth Actions Examples

- “late” in American SL: no mouth action

- “not yet” in American SL: the tongue touches the lower lip

Source: ASL University (http://www.lifeprint.com/)
Mouth Actions Examples

- “brother” in German SL: mouth frown

- “sister” in German SL: mouth stretch

Outline

- Is mouth linguistically significant for Sign Language?
- Future work, challenges and potentials.
Mouth Modeling in ASLR


Mouth Modeling in ASLR

Mouth Modeling in ASLR

- There is limited work on mouth modeling for the task of ASLR.

- We categorize the existing works with respect to:
  - Mouth modeling and tracking method
  - Mouth features
  - Recognition/Classification technique
  - Linguistic phenomena
  - SL
Mouth Modeling in ASLR

- Mouth modeling and tracking categorization:
  - Elliptical structure: [1] Parashar
  - Active Appearance Model: [2,3] v. Agris et al.
    [5,6] Schmidt et al.
    [8,9] Koller et al.
Mouth Modeling in ASLR

Mouth features categorization:

- Shape/Geometric measures: [2,3] v. Agris et al.  
  [5,6] Schmidt et al.

- Appearance: [1] Parashar  

- Both: [8,9] Koller et al.  
Mouth Modeling in ASLR

- Recognition/Classification categorization:
    [5,6] Schmidt et al.  
    [8,9] Koller et al.
Mouth Modeling in ASLR

- Linguistic phenomena categorization:
  - Negation, Questions, Conditional/Relative clause, Assertions, Sign boundaries: [1] Parashar
  [10] Benitez-Quiroz et al.
  
  - Mouthings: [5,6] Schmidt et al.
    [8,9] Koller et al.
Mouth Modeling in ASLR

- SL categorization:
  - American: [1] Parashar
    [10] Benitez-Quiroz et al.
  - German: [2,3] v. Agris et al.
    [5,6] Schmidt et al.
    [8,9] Koller et al.
- This is due to the existence of large annotated databases on these SLs.
Outline

- Is mouth linguistically significant for Sign Language?
- Future work, challenges and potentials.
Challenges and Potentials

- Automatic analysis of mouth non-manuals is a very challenging problem
  - Occlusion by hands, intense mouthings, expressions and pose, tongue visibility, low resolution of the mouth region

- It can be separated in two sub-problems:
  1) Automatic understanding of mouth-related expressions
  2) Automatic understanding of mouthings
Automatic understanding of mouth expressions

- It can greatly benefit from the extensive research on Automatic Analysis of Facial Expressions.

- It involves two main lines of research:
  - **Message judgment**
    Recognize the meaning (emotion) conveyed with a facial expression (e.g. six basic emotions).
  - **Sign judgment**
    Recognize the physiological manifestation of a facial expression into its fundamental and, arguably, irreducible atoms, such as the movement of individual facial muscles (e.g. FACS).
Automatic understanding of mouth expressions

- Message judgment is not suitable for ASLR
  - Discrete set of predefined messages (expressions) that does not cover the full range of possible SL expressions.
  - There is no universal set of predefined SL expressions.

- Sign judgment is relevant to ASLR
  - Every possible facial expression can be comprehensively described as a combination of AUs.
  - AUs annotation is a hard task.
Automatic understanding of mouthings

- It can greatly benefit from the extensive research on Visual Speech Recognition.
- A viseme is a generic facial image that can be used to describe a particular sound (equivalent of phoneme in spoken language).
- Visemes and phonemes do not have one-to-one correspondence.
- There is existing research on representing visual speech data using latent variables.
- Viseme recognition is very challenging even for humans (reported error rate about 50%).
Conclusions

- Development of ASLR systems has the potential to support millions of Deaf people, as well as help linguists understand better SLs.
- ASLR has mainly concentrated on manual features.
- Recent research has shown that non-manuals (especially the mouth) play an important role.
- Very few papers attempt the fusion of manual and non-manual cues.
- Mouth modeling for ASLR can greatly benefit from the existing research in Facial Expressions Recognition and Visual Speech Recognition.
Thank you for your attention!