

Spatial analysis of hand positions in French cued speech (LfPC)

Brigitte Bigi

Aix-Marseille Université, France

<https://doi.org/10.36505/TheLinguisticProceedings/2025/17/02/003/000689>

Abstract

Speech production relies on both acoustic and visual cues. While lip movements convey valuable information, many phonemes remain visually ambiguous. Cued Speech is a visual coding system that complements lipreading by combining mouth shapes with hand cues—specific handshapes and positions around the face—that encode consonants and vowels. It provides full visual access to spoken language and supports phonological awareness, literacy development, and oral language acquisition in deaf or hard-of-hearing individuals. This study investigates the spatial distribution of vowel-related hand positions in French Cued Speech (LfPC), which uses eight handshapes for consonants and five distinct positions for vowels.

Keywords: cued speech, deaf, hand position, vowels

Introduction

The French Cued Speech adaptation makes use of eight handshapes to encode consonants and five facial positions to encode vowels. Each syllable is represented by a combination of lip movement and a hand cue, also called a key, formed by a handshape–position pair. A simple syllable like CV or V is coded by a single key, while more complex structures, such as CCV, require multiple successive keys: for example, a 'C-' followed by a 'CV' structure (see Figure 1).

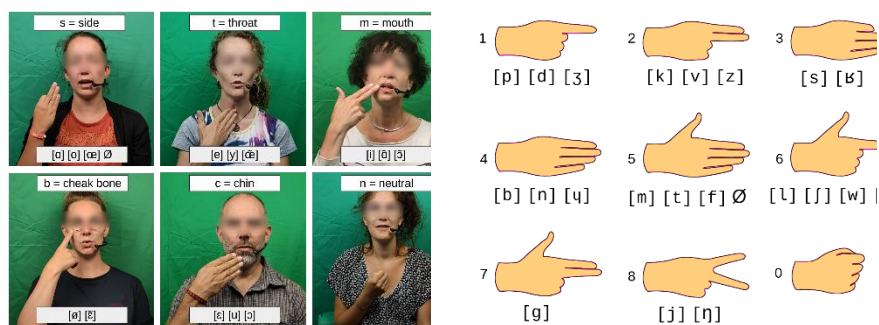


Figure 1. Positions representing vowels and handshapes representing consonants, in LfPC.

Below is a concrete example showing how a sentence is encoded into cues:

© The International Linguistic Society

Proceedings Linguistics 2025 Paris: 17th International Conference on Linguistic Research and Applications

text: Tu es gris.
phones: t y e ɣ ʁ i
sequence: C V V C C V
cues-structure: C-V.-V.C-.C-V
cues code: 5-t.5-t.7-s.3-m

In practice, producing CS requires the hand to simultaneously adopt the expected shape and reach the correct position near/on the face.

The present study focuses on the distributional statistics of (x, y) coordinates of keys at the moment the hand reaches its target position.

Methods and materials

This study is based on CLeLPC - *Corpus de Lecture en Langue française Parlée Complétée*, an open-access multimodal dataset of French Cued Speech (Bigi, 2022). Each participant read on a predefined topic containing four sessions: 32 CV syllables, 32 words or phrases, 7 to 10 isolated sentences, and one short text. For each participant, recordings were made under two distinct cueing instructions: 1/ a pedagogical condition (syllables and words/phrases reading), where participants were asked to produce cues as clearly as possible, 2/ a natural condition (sentences and texts reading), where they were instructed to cue as they would in spontaneous communication. The data for this study were drawn from a subset of the corpus comprising five highly proficient speakers.

Orthographic transcriptions were performed manually into the Inter-Pausal Units, normalized and converted into phonemes, then aligned with the audio signal using SPPAS (Bigi, 2015). All alignments were manually verified and corrected as necessary. Hand transitions were annotated frame-by-frame by a single expert. For each cued phoneme sequence, the frame at which the target position was reached was identified, allowing the extraction of spatial coordinates relative to the speaker's face. A total of 3,901 values were collected, normalized to a facial reference model (1000x1000 px), and analyzed to estimate mean positions and spatial variability.

Results

Table 1 & Figure 2 summarize vowel-related positions, associated phonemes, number of cues, and the mean and standard deviation of x and y coordinates. The four positions anchored on the face—cheekbone, chin, mouth, and throat—show relatively low variability, indicating stable production. In contrast, the 'side' position, which lacks a physical anchor point, displays the highest dispersion.

Figures 3a to 3d show results depending on the nature of the task, which influences spatial precision and variability. Cues produced during pedagogical

instruction (syllables and words) tend to be more stable than those produced during natural instruction (sentences and text).

Figures 4a to 4e show results depending on the speaker. While the overall structure of the system remains robust, individual strategies emerge, particularly in how some speakers realize peripheral positions such as “side” and “throat.”

Table 1. Distributional statistics of (x,y) coordinates for targeted positions.

Position	Phonemes	Number	Mean(x, y)	Stdev(x, y)
cheekbone	[ø], [ɛ̃]	230	(144, 397)	(70, 92)
chin	[ɛ], [ɔ], [u]	529	(425, 888)	(112, 76)
mouth	[i], [ā], [ɔ̃]	676	(260, 682)	(79, 91)
side	[a], [ə], [œ], [o], ∅	1750	(-193, 635)	(241, 233)
throat	[y], [e], [œ̃]	584	(435, 1419)	(143, 223)

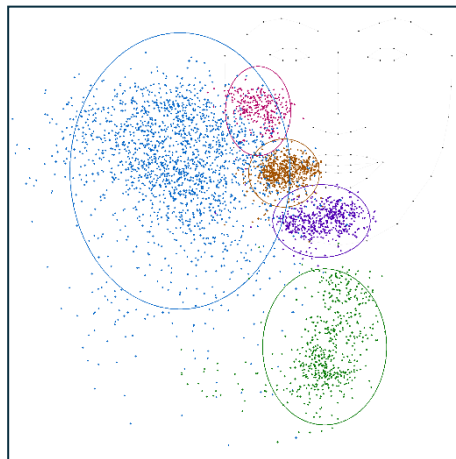


Figure 2. Visual distribution of hand targeted positions.

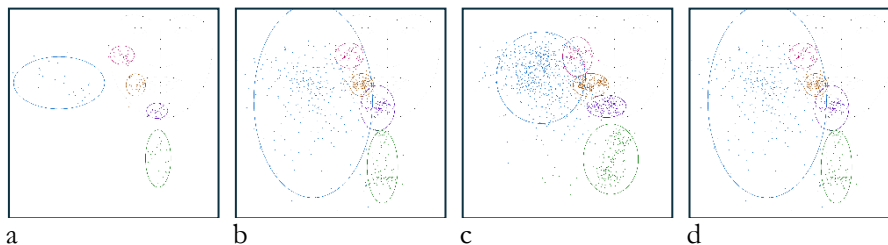


Figure 3. (a) CV syllables, (b) words, (c) sentences, and (d) text conditions.

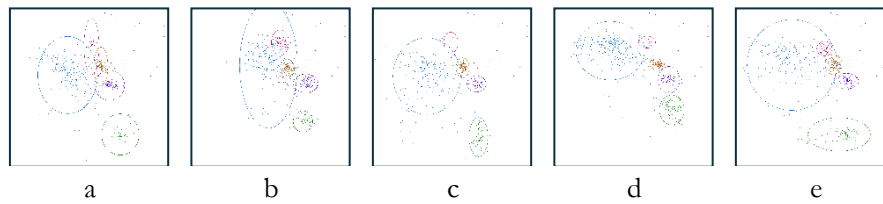


Figure 4. (a) CH, (b) VT, (c) AM, (d) ML, and (e) LM speakers

Conclusions

Despite extensive research on lipreading, few studies have quantitatively examined the spatial precision of hand cues in Cued Speech. In this study, mean coordinates are clearly differentiated across positions, confirming that vowel cues are spatially distinct and consistently produced. Further analysis revealed systematic variations depending on the nature of the task: natural reading versus pedagogical intent. This suggests that communicative purpose influences both precision and variability in cue production. Speaker-specific analysis confirmed the overall stability of the system, but also highlighted individual coding strategies, particularly for 'side' and 'throat' positions.

These results confirm the structured spatial organization of vowel cues in LfPC. Despite inter-speaker and task-related variation, the system ensures clear visual distinctions. The high dispersion of the 'side' cue emphasizes the role of anchoring in spatial accuracy. These findings inform linguistic modeling and practical applications, such as cue recognition or synthesis, and contribute to understanding the visual structure of Cued Speech. Future work will explore modeling of cue trajectories to support computer-based generation systems.

Acknowledgements

This research was supported by FIRAH (Fondation Internationale de la Recherche Appliquée sur le Handicap), project APa2022_022. Warm thanks also go to the members of the French ALPC association for their generous collaboration, their trust, and their invaluable insights throughout the development of this work.

References

- Bigi, B., Zimmermann, M., André, C. 2022. CLeLfPC: a Large Open Multi-Speaker Corpus of French Cued Speech. In Proceedings of the 13th Language Resources and Evaluation Conference, pp. 987-994, Marseille, France.
- Bigi, B. 2015. SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech. *The Phonetician*, 111–112, pp. 54–69.