



UMR 7309 - CNRS - AMU
PAROLE ET
LANGAGE



Aix-Marseille
université

CLeLPC: a Large Open Multi-Speaker Corpus of French Cued Speech



Brigitte Bigi, Maryvonne Zimmermann, Carine André

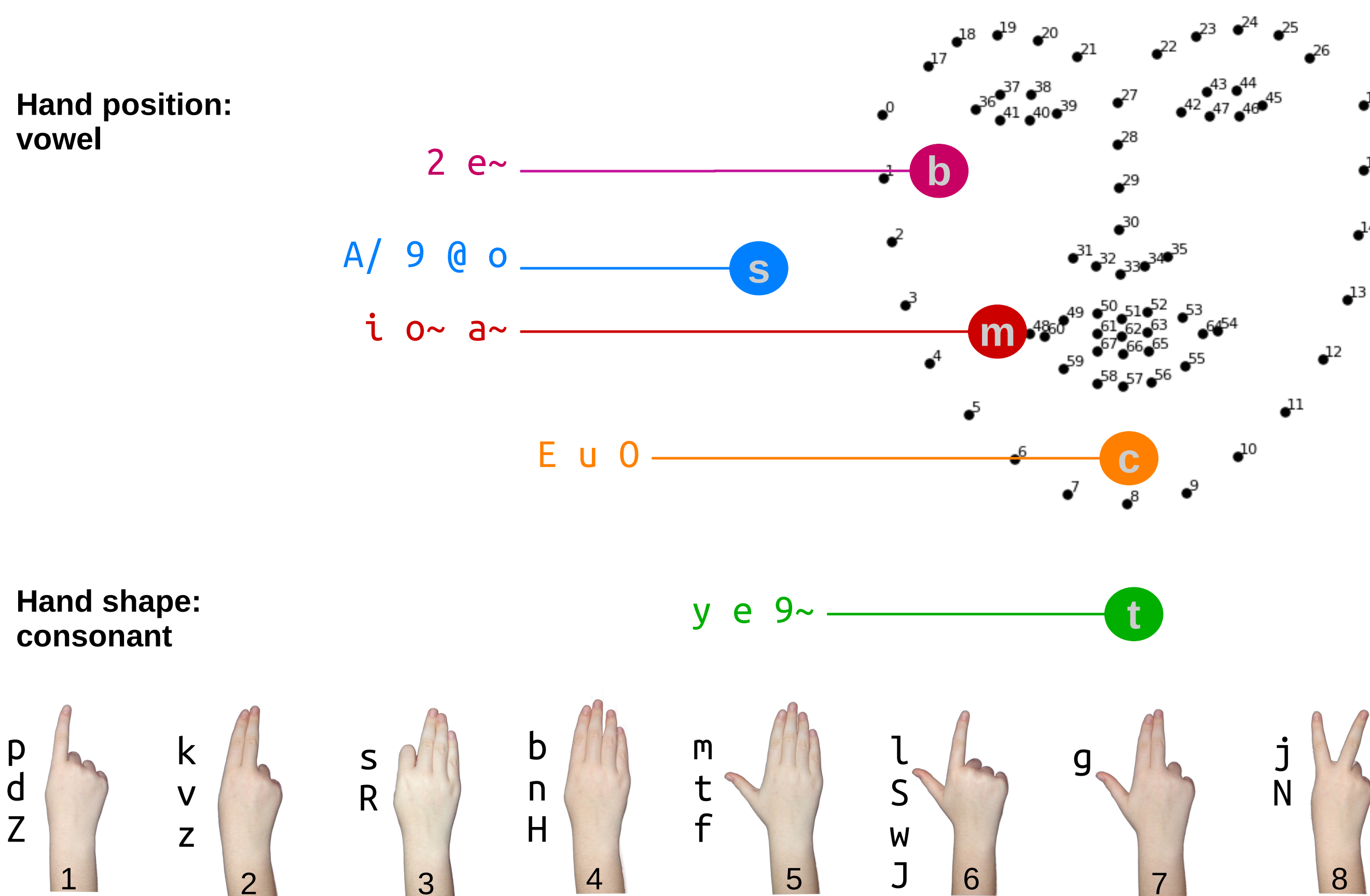
brigitte.bigi@cnrs.fr, maryvonne.zimmermann@datha.io, carine.andre@univ-amu.fr

ABSTRACT

CLeLPC - *Corpus de Lecture en Langue française Parlée Complétée*, is a corpus of French Cued Speech. It consists in about 4 hours of audio and HD video recordings of 23 participants. The recordings are 160 different isolated CV syllables repeated 5 times, 320 words or phrases repeated 2-3 times and about 350 sentences repeated 2-3 times. The corpus includes orthographic transliteration and other phonetic annotations on 5 of the recorded topics. The corpus is licensed under CC-BY-NC-4.0 International License. It can be used for any further research or teaching purpose.

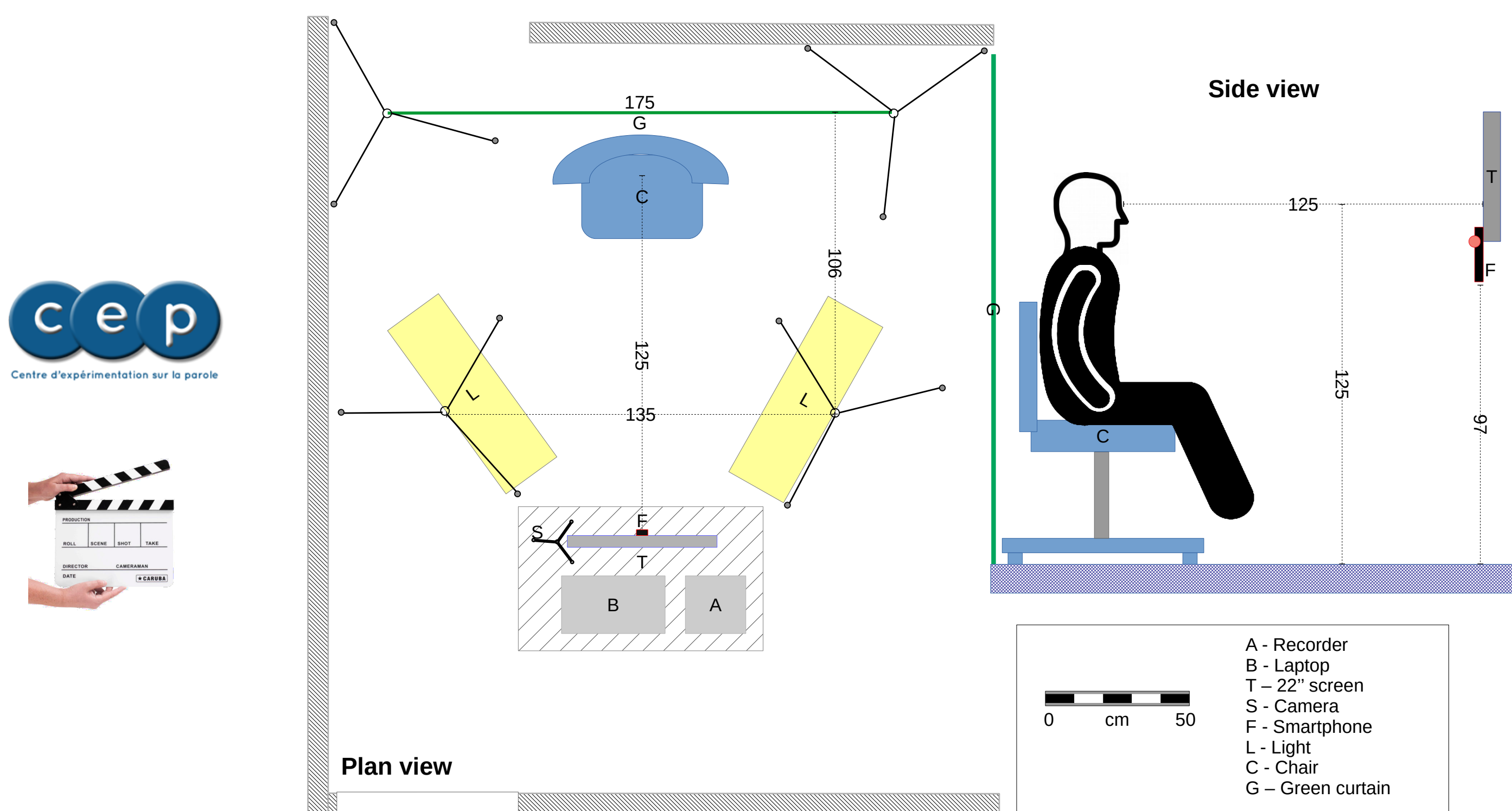
CUED SPEECH

R. Orin Cornett invented the Cued Speech in 1966. It's a visual communication mode developed for **deaf people** to complement *speechreading at the phonetic level with hands*. It uses handshapes in different placements near the face in combination with the mouth movements of speech to make *the phonemes look different from each other*. Ones sounds are made visible and look different, it results in a better understanding of spoken language. A single CV syllable is generated or decoded through both: the lips position and the key of the hand:



Cued Speech for French language (LfPC)

CORPUS RECORDING PROTOCOL



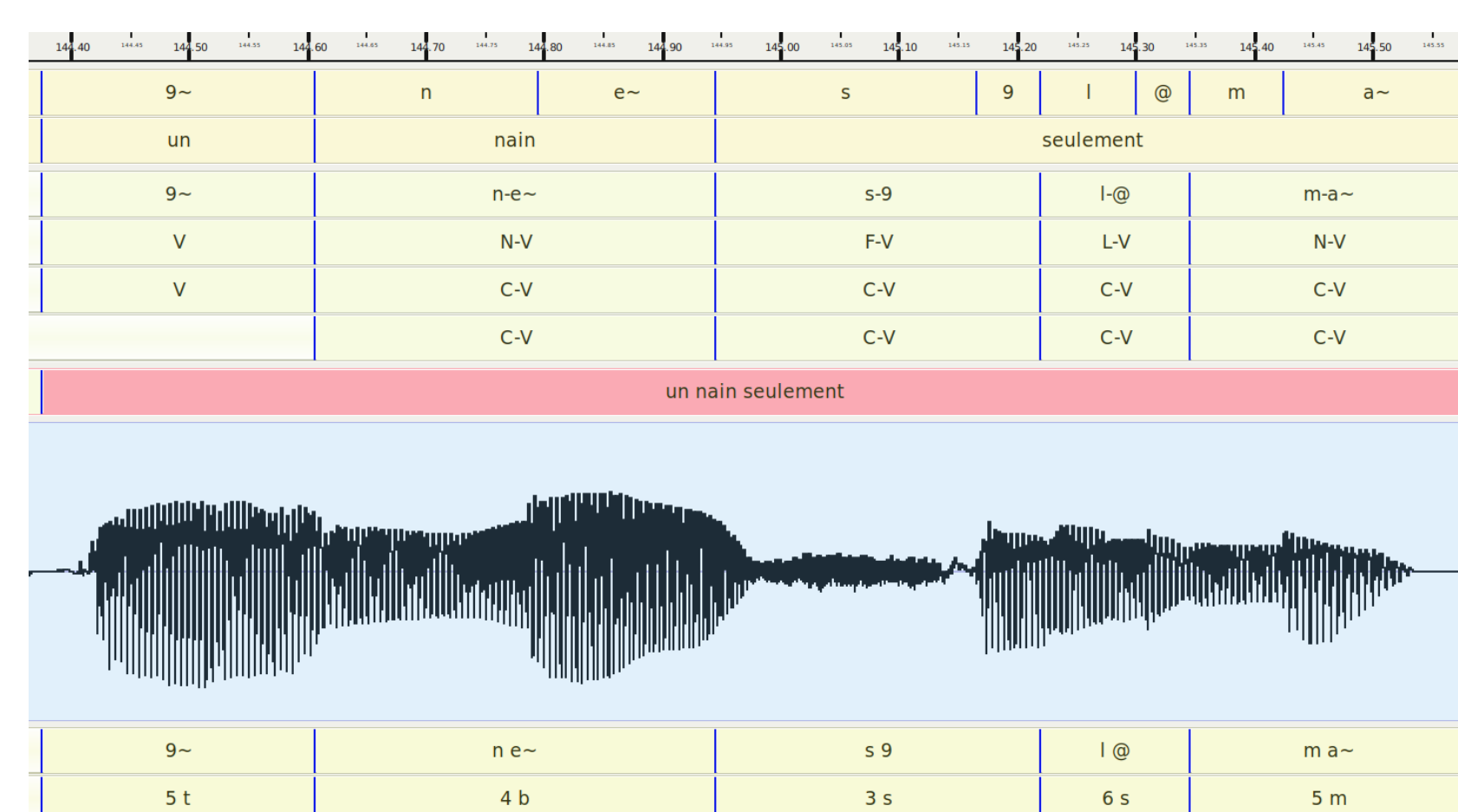
- recorded in a calm hotel room
- audio: headworn cardioid microphone AKG C520
- audio: Zoom LiveTrak L-8 recorder, powered by an external battery charger
- front camera: a smartphone Xiaomi 10 lite 5G with a 1080p, 1080x1920, 60fps, MP4
- side camera: a canon Legria G5, 1920x1080, 25 fps, MPEG
- we developped a program to synchronize automatically each of the video stream with the corresponding audio from the clap time values (open source)

CORPUS DESCRIPTION

- 4 hours of audio/video recordings
- 23 recorded participants (5 men - 18 women, 25-59 years old)
- 2 reading instructions: (i1) read clearly, like to teach CS to someone else; (i2) read naturally, like to tell or read someone a story
- 10 different topics: <http://www.sppas.fr/LFPC/>
- 25 recorded topics. Each topic is made of 4 sessions:
 - 32 isolated "CV" syllables (i1)
 - 32 isolated words or phrases (i1)
 - isolated sentences (i2)
 - a text divided into 4-7 parts (i2)
- hosted by Ortolang: <https://hdl.handle.net/11403/clelfpc>
- publicly available for academic members



- the 4 sessions of 5 topics are annotated
- Manually checked annotations: IPU, orthographic transcription, time-aligned phonemes
- Automatic annotations (SPPAS): syllables, expected keys



- PhonAlign
- TokensAlign
- Syllables
- SyllClasses
- SyllStructures
- FilteredSyllStructs
- TranscriptionIPUs
- CuedSpeechSyll
- CuedSpeechKeys

PRELIMINARY RESULTS

Only estimated on the 2156 syllables with 'CV' structures

	Occ.	Mean	StDev
(Attina, 2005), SC subject	159	0.253	0.045
(Aboutabit, 2007)	57	0.284	0.075
5 'CV' syll' sessions - i1	159	0.354	0.105
5 'word' sessions - i1	458	0.320	0.101
5 'sent' sessions - i2	741	0.271	0.083
5 'text' sessions - i2	798	0.253	0.084

Key	Occ.	Mean	Key	Occ.	Mean
(1)+(b)	57	0.319	(5)+(b)	17	0.281
(1)+(s)	346	0.216	(5)+(s)	416	0.199
(1)+(m)	71	0.300	(5)+(m)	200	0.242
(1)+(c)	72	0.256	(5)+(c)	138	0.262
(1)+(t)	119	0.304	(5)+(t)	242	0.235
(2)+(b)	12	0.255	(6)+(b)	20	0.304
(2)+(s)	253	0.221	(6)+(s)	353	0.205
(2)+(m)	88	0.272	(6)+(m)	64	0.274
(2)+(c)	86	0.274	(6)+(c)	80	0.295
(2)+(t)	55	0.303	(6)+(t)	88	0.264
(3)+(b)	23	0.285	(7)+(b)	1	0.400
(3)+(s)	489	0.196	(7)+(s)	55	0.210
(3)+(m)	131	0.301	(7)+(m)	11	0.350
(3)+(c)	82	0.280	(7)+(c)	7	0.327
(3)+(t)	110	0.302	(7)+(t)	11	0.311
(4)+(b)	5	0.305	(8)+(b)	14	0.223
(4)+(s)	164	0.196	(8)+(s)	58	0.231
(4)+(m)	84	0.270	(8)+(m)	20	0.236
(4)+(c)	38	0.283	(8)+(c)	20	0.228
(4)+(t)	53	0.299	(8)+(t)	49	0.284

Number, mean duration (in seconds) and standard deviation of the 'CV' syllables compared to previous results estimated on the 3rd syllable of sequences of 4 syllables.

Number and mean duration of expected keys

Main results:

By comparing mean durations among the different shapes and among the different hand positions, it seems that:

- the hand position has a high influence on the key duration;
- the hand shape has not.