Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 987–994
Marseille, 20-25 June 2022

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

Brigitte Bigi(1), **Maryvonne Zimmermann**(2,3), **Carine André**(1)
(1) LPL, CNRS, Aix-Marseille Univ, (2) ALPC, (3) Datha
(1)13100 Aix-en-Provence, (2)75015 Paris, (3)94400 Vitry-sur-Seine
brigitte.bigi@cnrs.fr, maryvonne.zimmermann@datha.io, carine.andre@univ-amu.fr

Abstract

Cued Speech is a communication system developed for deaf people to complement speechreading at the phonetic level with hands. This visual communication mode uses handshapes in different placements near the face in combination with the mouth movements of speech to make the phonemes of spoken language look different from each other. This paper describes CLeLfPC - Corpus de Lecture en Langue française Parlée Complétée, 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 is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It can be used for any further research or teaching purpose. The corpus includes orthographic transliteration and other phonetic annotations on 5 of the recorded topics, i.e. syllables, words, isolated sentences and a text. The early results are encouraging: it seems that 1/ the hand position has a high influence on the key audio duration; and 2/ the hand shape has not.

Keywords: Cued Speech, corpus, multimodality

1. Introduction

The production of speech naturally involves lip movements; both the acoustic information as well as the lipreading are part of the phonological representation of hearing people. The processing of the audible acoustic information is then influenced by the visual information (McGurk and MacDonald, 1976). For a better comprehension every sound of the language should look different but many sounds look alike on the lips when speaking. As a consequence, a large number of words have the same lip movements and can't be distinguished. The accuracy for lipreading sentences rarely exceeds 10%-30% words correct (Rönnberg, 1995; Rönnberg et al., 1998).

In 1966, R. Orin Cornett invented the Cued Speech (Cornett, 1967), a visual system of communication; it adds information about the pronounced sounds that are not visible on the lips. Cued speech is a code to represent each sound of a given language with a shape of the hand for a consonant, and a position around the face for a vowel. Actually, from both the hand position on the face and hand shapes, CV syllables are represented. So, a single CV syllable will be generated or decoded through both the lips position and the key of the hand. Each time a speaker pronounces a 'CV' or 'V' syllable, a cue is produced and other syllabic structures are produced with several cues - for example, a 'CCV' syllable is coded with the two keys 'C' then 'CV'. When sounds look alike on the lips, they are cued differently. Thanks to this code, speech reading is encouraged since the Cued Speech keys match all of the spoken phonemes but phonemes with the same movement have different keys. Ones sounds are made visible and look different, it results in a better understanding of spoken language.

The World Health Organization reported that more than 5% of the world's population has a hearing loss, i.e 432 million adults and 34 million children. It is estimated that by 2050 over 700 million people – or one in every ten people – will have disabling hearing loss [1]. Several studies have been conducted on Cued Speech (CS) to show how it can help speech perception for deaf or hard of hearing persons. Cued Speech improves speech perception for hearing-impaired people and it offers a complete representation of the phonological system for hearing-impaired people. CS has been adapted for more than 60 languages[2] and is increasingly popular. CS improves the speech reading capabilities of profoundly deaf 8-to 12-year-old subjects(Clarke and Ling, 1976). CS is of great help by improving communication between deaf or hard of hearing children and their hearing family members. An added benefit of being cued to is that it is unconsciously training the deaf child to develop exceptionally good lip-reading skills that they can then use to understand people who are not cuing(Neef and Iwata, 1985). CS is also significantly improving speech reading abilities of prelingually deaf persons (Kaplan, 1975).

The use of hearing Assistive Technology[3] can further improve access to communication and education for people with hearing loss. Cued Speech can enhance the benefits of cochlear implants by training the brain to make better use of the signal from the cochlear implant (Leybaert et al., 2010): "an exposure to Cued Speech before or after the implantation could be im-

[1] https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss visited: 11-2021

[2] https://www.academieinternationale.org/list-of-cued-languages visited 11-2021

[3] Assistive Technology: products, equipment, and systems that enhance learning, working, and daily living for persons with disabilities. https://www.atia.org/

portant in the aural rehabilitation process of cochlear implantees".

Despite the huge number of studies demonstrating the benefits of cued speech, studies on the automatic recognition or generation of CS are rather rare. (Duchnowski et al., 1998) assessed the feasibility of automatic determination and presentation of cues and designed a prototype of a real-time automatic cuing system. In this system, a speaker is filmed speaking without coding and in another room, the image of the speaker with the synthesis keys according to the rules of the Cued Speech is displayed on a screen. Several versions of the system were evaluated and it resulted in at least a small benefit to the cue receiver relative to speechreading alone. However, they didn't investigated the motor organisation of Cued Speech production, i.e. the coarticulation of Cued Speech articulators.

It was already found in (Cornett, 1967) that lips and hand movements are asynchronous in CS. A study of the temporal organisation of the acoustic indices in relation to the movement of the lips and the hand shape was investigated on French language in (Cathiard et al., 2003) and continued in (Attina, 2005; Aboutabit, 2007) with the proposal of a synchronization model. Lips movement is more related to the phoneme production and hand movement is more related to the speech syllabic cycle, and that the handshape began to be formed a long time before the acoustic consonant. (Liu et al., 2018) proposed a novel hand preceding model to predict the temporal segmentation of hand movements only from the audio based segmentation in CS. In (Liu et al., 2019), authors investigate and confirm the phenomenon of hand preceding lips in British English CS and compare their results with their ones on French CS.

To support all these studies, several corpora were created and two of them were made available. The French corpus is made of video recordings of a CS speaker upper body with a 720x576 images resolution at 50 fps, and audio files recorded in a sound-proof booth (Liu, L. and Hueber, T. and Feng, G. and Beautemps, D., 2018). It is a set of 238 French sentences repeated twice by the speaker. The distributed part of the corpus contains the videos, the audio, the prompts and the automatically time-aligned phonemes. The British English Cued Speech dataset was recorded in Cued Speech UK association (Liu, L., 2019). The speaker was cuing a set of 100 British English sentences. Video of the interpreter's upper body are recorded at 25 fps, with a resolution of 720x1280. The distributed part of the corpus contains the MP4 video files and the prompts. Four new speakers were added for a study on hand shape recognition and phoneme recognition described in (Wang et al., 2021). In both French and English corpora, the speakers are certified in transliteration speech into CS.

This paper describes CLeLfPC, a large open source multi speaker dataset of Cued Speech (Bigi, B. and Zimmermann, M., 2021). The corpus was recorded during an event organized by the ALPC, the National Association for the Promotion and Development of the French Cued Speech, in 2021 August. Among others, this new CS corpus brings the following tangible benefits: an HD video quality of the whole speaker, not only the upper body in order to capture the whole hand movements; 23 different participants, some are certified and some are not; 160 different isolated 'CV' syllables repeated 5 times by different participants; 320 different isolated words or phrases repeated 2 or 3 times by different participants; about 350 sentences repeated 2 or 3 times by different participants; all the annotations of the corpus was, are and will be made available within new versions of the corpus and under the terms of the CC-BY-NC, the Creative Commons Attribution-Non-Commercial 4.0 International License. This corpus of 4 hours of audio/video recordings can be used for any further research or teaching purpose about CS.

2. The French Cued Speech

The term "Langue française Parlée Complétée" (LfPC) is the French language term for French CS. It literally means "Supplemented Spoken French Language". The phonemes of the 8 LfPC shapes of the hand are noted in X-SAMPA as follow:

(1) /p/, /d/, /Z/

(2) /k/, /v/, /z/

(3) /s/, /R/

(4) /b/, /n/, /H/

(5) /m/, /t/, /f/, no consonant

(6) /l/, /S/, /J/, /w/

(7) /g/

(8) /j/, /N/

The 5 LfPC hand placements of vowels are:

(b) cheek bone /e∼/, /2/

(s) side /a/, /o/, /9/, /@/, no vowel

(m) mouth /i/, /O∼/, /a∼/

(c) chin /E/, /u/, /O/

(t) throat /y/, /e/, /9∼/

In order to illustrate both CLeLfPC and the French CS system, we extracted several images of the front-video (Figure 1). They represent the 8 hand shapes and 5 hand placements of French CS by 8 participants: 6 are cuing with the right hand and 2 with the left one, 6M/2F, 2 are impaired.

Figure 1: French CS keys represented with screenshots of CLeLfPC

3. Corpus recording protocol

3.1. Read speech

We have prepared a set of 10 topics. We asked participants to read aloud and to cue one topic; two participants accepted to read 2 different topics. Each topic was made of 4 sessions and the sessions were recorded separately for the participant to have a short break:

- 32 isolated ”CV” syllables;
- 32 isolated words or phrases;
- isolated sentences;
- a text divided into 4-7 parts.

The topics/sessions were carefully designed in order to cover a large amount of different keys and sequences of keys. Technically, they were organized in an HTML page with some javascripts in order to show them like slides. When the participant finished to read a slide, we pressed manually a ”next” button to display the next slide. This HTML page is available at the URL: `http://www.sppas.org/LFPC/`. This web page is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

3.2. Syllables selection method

The syllables selection aimed for the broadest possible phonetic-and-keys coverage. We selected 160 distinct CV syllables: they were spread across 5 different lists. These lists were assigned respectively to topics 1-5 and 6-10. We attempted to distribute consonants and vowels as equally as possible; for examples the 26 syllables of the hand shape (1) are distributed with 5 of them in each list plus one list with the remaining one, or the 21 syllables of the shape (2) are 3 up to 6 in each list. Because the participants were not familiar with phoneme symbols, we had to write them with a standard orthography in the slides; so we can’t be sure they read them exactly like we expected.

3.3. Words/phrases selection method

Like for syllables, words and phrases were also carefully selected. At a first stage, we choose them depending on the corresponding list of syllables. For a given topic, each CV syllable was included into a word or a phrase - that’s the reason why there are 32 words or phrases in each topic. We took care to add it with a left- and right- context, i.e. a syllable before and a syllable after. We also took care to select 2 different left-right contexts in the two topics that are using the same list of

syllables. This constraint will allow to compare the exact position of the vowel for the same key of the same syllable in the two conditions: isolated vs with contexts. Contrariwise to (Attina, 2005), we didn't paid attention to the previous and next syllable structure, so it's not necessarily a 'CV' one, like the word "Cognac" /koNak/. We finally paid attention to select words or phrases in order to obtain a broad coverage of key sequences. Obviously, some sound sequences are much more frequent in French than others and so does in our selection. We estimated that the selected words and phrases of a topic should result to about 150 up to 175 keys in each topic.

3.4. Sentences and texts selection method

We selected a set of un-licensed or public sentences and texts, including some texts from previous studies in the Phonetics domain. We also wrote some texts and the major part of the sentences. We attempted to make them as pleasant as possible to read because coding while reading is already a difficult task and we didn't wanted to add more difficulties than required.

3.5. Recording conditions

The recordings of a topic started by asking the participant to read instructions of the web page then to fill in an information sheet. It includes personal information - firstname, lastname, address, contact, left- or right-main hand. It then contains information about the participant knowledge of cued speech: coding hand, coding level, coding use, where and when cued speech was acquired, hearing impairment. It finally includes a checkbox to give us the authorization to record them with both audio and video. We then explained the recording conditions and gave the following reading instructions:

i1 the syllables and the words/phases have to be read clearly, like to teach CS to someone else;

i2 the sentences and the text should be read as naturally as possible, like to tell or read someone a story.

We adjusted the seat height. The microphone and the audio gain of the recorder was adjusted with a test sentence. We then recorded the 4 sessions, one at a time. Each session started by an announcement and a video "clap". Finally, the participant had to sign a consent form. The form contains a list of checkboxes in order to accept some use of the recordings. All participants agreed the following uses:

- for any scientific research;
- for educational purposes;
- to share in research projects;
- to share with the ALPC.

4. Recording equipment and place

This section describes the equipment we used and the place we recorded participants. In the scope of making this corpus reproducible, technical information are of great importance.
Our choices of equipment were determined by the following constraints: the equipment we already had and the budget to purchase new ones. The size, weight and bulkiness were also important particularly because we didn't knew anything in advance about the recording place.

4.1. Place of recording

We recorded the corpus in a calm hotel room with beds set aside; the available room to install the equipment was 1.80m x 2.50m. Figure 2 is a foot plan and a side plan representing the recording place.

4.2. Equipment

A 22" LCD screen of 52cm width was used to display the HTML page. Mozilla web browser was used; the text-size was ranging 100-120% depending on the participant. The slides were displayed at an average of 125cm height. In front of the screen, we installed a seat that was locked in order to move only in the top-bottom way. We adjusted it for each participant: the height of the eyes had to be at about 125cm. Behind the seat, we set up a 175cm (width) x 180cm (height) green screen background on a crossbar maintained by two backdrop support stands to solve the problems of shadow, reflective point, color interference, etc. We also installed two led lights at left and right of the screen, a little bit higher than it to avoid blinding the participant - height of the support stands was 180cm.

4.2.1. Audio

We carefully tested and compared several recording systems and microphones. We then selected the headworn cardioid microphone AKG C520, with a Phantom power adapter of 48V connected with an XLR to the recorder. The foam windscreen that comes with the microphone was not used in order to not hide the mouth on the video; but it results in a whisper in the recorded audio files when participants are breathing.
The audio was recorded with a Zoom LiveTrak L-8 which was powered by an external battery charger. A separate track is recorded for each microphone. The track is a one-channel Waveform Audio File Format with .WAV extension in 16 bits, 48,000Hz.
Three microphones were connected to the recorder. A first one was installed at the left of the headworn and was used only for participants cuing with their right hand. A second one was installed at the right of the headworn and was used only for participants cuing with their left hand. A third one was used for the session announcement only which is useful when processing master recordings in order to create the corpus with time-synchronized files.

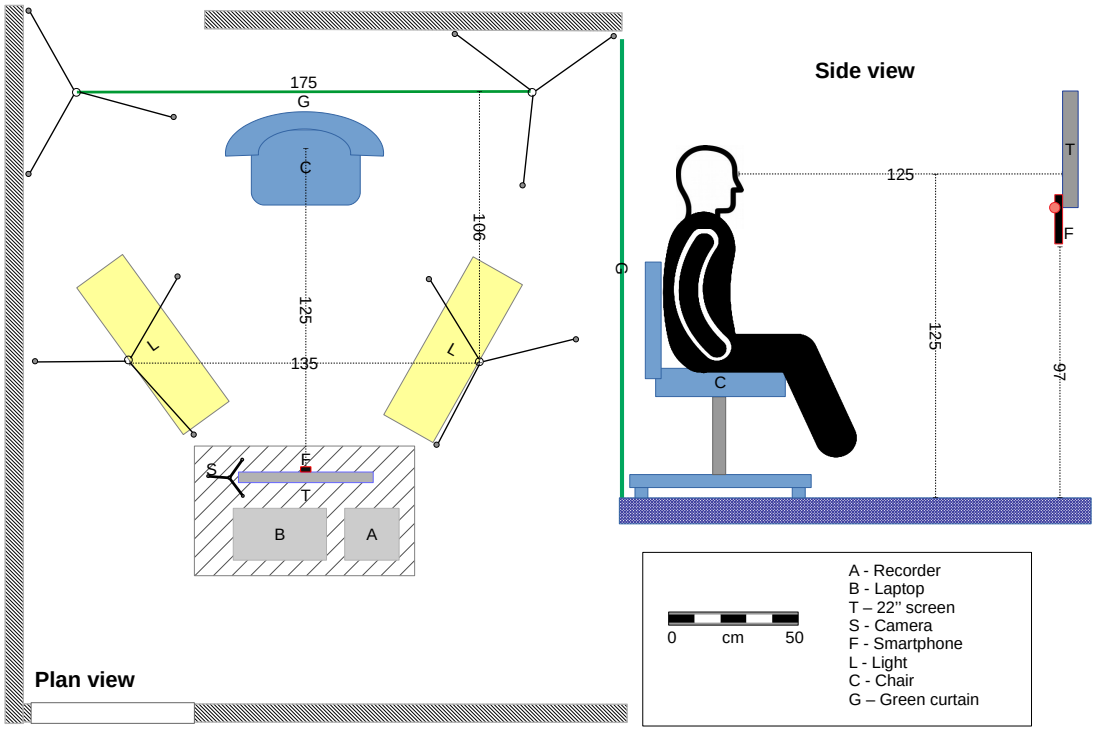

Figure 2: Recording room design

4.2.2. Video 1 - front

A front camera was placed at the bottom of the screen, exactly at the middle. We used a smartphone Xiaomi 10 lite 5G with a 1080p, 60fps camera. With this equipment, videos are embedded in the MPEG-4 container (.mp4 file extension) with libx264 encoding. We selected an image size of 1080x1920, i.e. 9:16 rate. It has to be noticed that unlike a real camera, the 60 fps are an average. We observed in the recorded video streams that it can range from 59.98 to 60.02. The recorded audio stream of the smartphone was an ac3 format with stereo-16bits-48,000Hz of very poor quality.

4.2.3. Video 2 - side

At the side of the screen, a canon Legria G5 camera was installed. It recorded videos with 1920x1080 image size - i.e. 16:9, at 25 fps, with the MPEG transport stream container (.MTS file extension); the video stream was HEVC encoded. The recorded audio stream of the camera was an aac format with stereo-16bits-48,000Hz of acceptable quality.

This camera was installed at right of the screen to record participants cuing with the right hand and at left of the screen to record participants cuing with their left hand.

5. Recorded participants and files

5.1. The participants

Every year, the ALPC organizes training programs designed to improve the qualifications of participants in coding LfPC. The corpus was then recorded during the summer 2021 session, in August 24-26 at "Les Karellis" (Savoie, France)[4].

All 23 recorded participants have voluntereed. There were 25-59 years old - average is 40; there are 5 men and 18 women. They were participating at the event because they are either people with hearing loss, or a family member of someone with hearing loss, or PhD students working on CS, or CS professionals.

5.2. From the master recordings to the corpus

When several different equipment contributed to the data, the first step consists in synchronizing the master recording files. For CLeLfPC, we had 3 files - the main audio, the front video and the side video, with 5 streams (3 audios + 2 videos).

At a first stage, it was required to extract the audio from the 2 video files. We developed a program in Python language which is a wrapper for the 3 following open source programs: ffmpeg[5], sox[6] and SPPAS[7]. This program "montage.py" is distributed under the terms of the GNU GPL v3.0 or later license.

[4]For details about the event, see https://alpc.asso.fr/stage-2021-quarantiemes-rugissants/

[5]https://ffmpeg.org/

[6]http://sox.sourceforge.net/

[7]http://www.sppas.org/

The program synchronizes each of the video stream with the corresponding audio. It requires a spreadsheet file containing the following columns:

1. Speaker identifier (2 chars);
2. Recorded topic number ranging 1-10 (2 chars);
3. Coding hand (1 char): 'g' for left, 'd' for right;
4. Main hand (1 char);
5. French cued speech level ranging 1-6 (1 char);
6. Frequency of use ranging 1-4 (1 char);
7. Hearing impairment: 0 for no, 1 for yes (1 char);
8. Gender: m/f (1 char);
9. Session name: syll/word/sent/text (4 chars);
10. Audio filename;
11. Video side filename;
12. Video front filename;
13. Audio clap time;
14. Video side clap time;
15. Video front clap time;
16. Delta clap, it's a duration to be applied before (neg value) or after (pos value) the clap time;
17. Duration to keep.

This file had to be filled in manually. The information of the first 9 columns are extracted from the information sheet and the recording context. They are used to create the output filename; an example is "syll_2_MZ_dd520f" which means that the recorded session is the syllables of topic number 2, the participant is "MZ" and her main hand is the right one, her coding hand is the right one, her cued level is 5, her frequency of use is 2 and she has no hearing impairment. The other columns are related to the recorded files and the timing. The clap time values were all identified from the audio files in Audacity software tool, by zooming the clap in order to have a precision of about 1-3 milliseconds.
The program is synchronizing the audio with each one of the video streams with the following algorithm:

- evaluate the time value `bt` and the frame `bf` in which the clap occurs in the video and add delta;
- evaluate the time value `et` and the frame `ef` in which the video is ending, and add delta+duration;
- add silence or trim the beginning of the audio file to correspond to the one of the video;
- add silence or trim the end of the audio to match the video duration;
- trim the audio from `bt` to `et`;
- trim the video from `bf` to `ef`.

The original audio format was preserved: it remained a WAV, 16bits, mono, 48000Hz. Each video stream was decoded in order to trim it and then re-encoded with the High Efficiency Video Coding (H.265) using the libx265 library of ffmpeg. We choose a very small compression rate of 14. To store the video stream, we selected the Matroska Multimedia Container because it is a powerful free and open container format (.mkv). Both synchronized audio and video files have then exactly the same duration and will be the main files for futher analyses. For conveniency, the program created a ".mp4" file with a lossy compression of audio/video streams.

6. CLeLfPC repository

The corpus is hosted by the French institutional repository Ortolang[8]. The following files are available to everyone: a demo video, the corpus description, the unfilled versions of the 2 consent forms and the python script to synchronize the recordings. In version 1, the 25 syllable sessions (synchronized audio/video files) were made available to any academic member (see the Ortolang policy for details about this). In version 2, the synchronized audio and videos of all the words/phrases sessions were made available to academic members; and the prompts of all sessions were added and available to anyone who has an account on the repository. In version 3, the annotations described in the next section were added.
The non-academic members have to write to the authors in order to get an open access to the corpus because the authors have to check if the corpus is requested for a research or a teaching purpose.

7. Annotations and early results

In the scope of preliminary analyses, we selected 5 different topics recorded by participants of level 5 or 6. We took care to cover the whole set of the 160 distinct 'CV' syllables: topic 1 (CH_dd640f), topic 2 (VT_dd640f), topic 3 (AM_dd630f), topic 9 (LM_gd640f) and topic 5 (ML_gg540f).
The 4 sessions of all the 5 topics were time-aligned at the phonetic level. Using SPPAS (Bigi, 2015), Inter-Pausal Units - e.g. sounding segments separated by silences, were identified (Bigi and Priego-Valverde, 2019). The orthographic transcription was then performed manually with Praat (Weenink, 1992 2021) by the first author of this paper, and the boundaries of the IPUs were manually verified at the same time. The text transcription was automatically normalized and converted to phonemes, the phonemes were manually verified then automatically aligned with the recording. For these purposes, we modified the linguistic resources of SPPAS in order to replace the existing meta-phonemes `O/` and `U~/` respectively by `o` or `O` and `e~` or `9~`; we also added `N` and `J` into the acoustic model and we modified the pronunciation dictionary. Finally, the time-aligned phonemes were manually verified with Praat by the first author. Figure 3 is illustrating the audio waveform of an IPU of the corpus with all the annotations.
In the previous works (Attina, 2005; Aboutabit, 2007), the synchronization model of hand-lips-syllable was

[8] `https://www.ortolang.fr/`

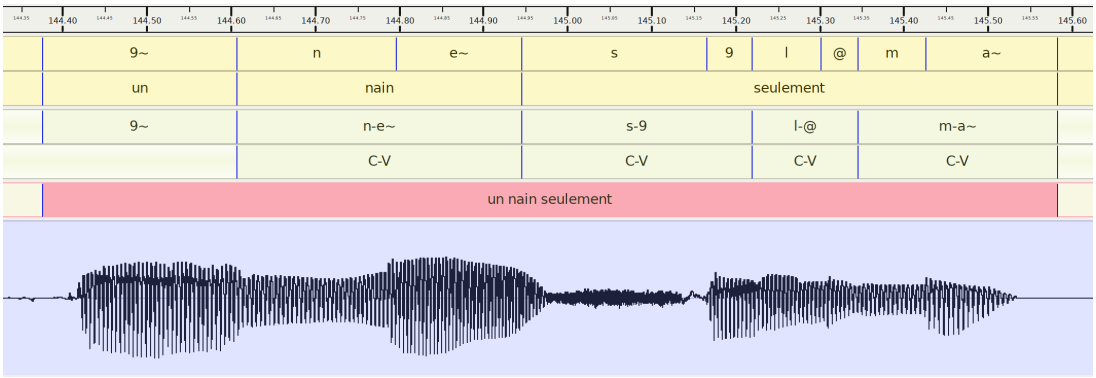

Figure 3: Waveform and annotations of an IPU of the corpus. Form top to bottom: phonemes, tokens, syllables, filtered 'CV' syllable structures, orthographic transcription into IPUs.

established by analyzing 'CV' syllables only. In the scope of getting a comparable result, we also analyzed only such syllables. SPPAS allowed to get the time-aligned syllables and their structures (Bigi et al., 2010) and its filtering system (Bigi and Saubesty, 2015) was used in order to select only the syllables with 'CV' structures. It also allowed to estimate distributionnal statistics: mean duration and standard deviation are reported in Table 1. As expected, the mean duration is significantly higher in sessions with the instruction **i1** compared to the instruction **i2**. This latter is very close to the ones reported in the previous works. However, we observe in CLeLfPC a significantly higher standard deviation. This is probably due to the fact that both there are 5 different speakers and the sentences/texts were more authentic (horoscope, weather report, cooking recipe, children's story...). We also measured that the consonant is representing 49.73% of the duration of the CV syllables (standard deviation is 14).

	Occ.	Mean	StDev
(Attina, 2005), SC subject	159	0.253	0.045
(Aboutabit, 2007)	57	0.284	0.075
5 '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

Table 1: Number, mean duration and standard deviation of the 2156 'CV' syllables (in seconds). They are compared to previous results which were estimated on the third syllable only of sequences of 4 syllables.

The phonemes were clusterized automatically into sequences of keys. This program was included into SPPAS (version 4.1) in the set of automatic annotations, with name "LPC key code". The algorithm is very close to the automatic syllabification, except that the key structures are only CV, V or C. We did not checked manually the video in order to compare the realized keys with these expected ones. Table 2 indicates the number and mean duration of these expected keys; the standard deviation is ranging 0.050-0.120. Except for (8), we observe that whatever the hand shape, the higher frequency of the hand position (s), the lower mean duration. For the hand shapes (4)-(5)-(6)-(7), the lower frequency of the hand position (b), the lower mean duration. By comparing mean durations among the different shapes and among the different hand positions, *it seems that 1/ the hand position has a high influence on the key duration; and 2/ the hand shape has not.*

However, these results have to be carefully interpretated: the phoneme segmentation should be checked by a second expert, the expected keys have to be compared to the realized ones, the 20 other sessions have to be annotated.

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

Table 2: Number and mean duration of expected keys

8. Conclusion

The appropriate use of hearing devices like hearing aids or cochlear implants, Assistive Technologies and social support can facilitate access to communication, education and equal opportunities to deaf children. An automatic generation system of CS could be a valuable one, and creating a CS corpus is the first required step toward such system. Keeping in mind this ambitious long-term project, this paper described CLeLfPC - Corpus de Lecture en Langue française Parlée Complétée, a corpus of French Cued Speech. It is made of 4 hours of audio/video recordings and partly annotated. Research is ongoing, but early results on these annotations are encouraging. The corpus is under the terms of the CC-BY-NC-4.0 and can be used for research or teaching purpose on Cued Speech.

9. Acknowledgements

We are grateful for the assistance and support of the CEP - Centre d'Expérimentation sur la Parole at the LPL. We address our special thanks to the ALPC for its support. We want to express our gratitude to the participants who all voluntereed and will allow us for doing the research we need for our project.

10. Bibliographical References

Aboutabit, N. (2007). *Reconnaissance de la Langue Française Parlée Complété (LPC): décodage phonétique des gestes main-lèvres.* Ph.D. thesis, Institut National Polytechnique de Grenoble - INPG.

Attina, V. (2005). *La Langue Française Parlée Complétée: Production et Perception*. Ph.D. thesis, Institut National Polytechnique de Grenoble - INPG.

Bigi, B. and Priego-Valverde, B. (2019). Search for inter-pausal units: application to cheese! corpus. In *9th Language & Technology Conference*, pages 289–293, Poznań, Poland.

Bigi, B. and Saubesty, J. (2015). Searching and retrieving multi-levels annotated data. In *Proceedings of Gesture and Speech in Interactioni - 4th edition*, pages 31–36, Nantes, France.

Bigi, B., Meunier, C., Nesterenko, I., and Bertrand, R. (2010). Automatic detection of syllable boundaries in spontaneous speech. In *Language Resource and Evaluation Conference*, pages 3285–3292, La Valetta, Malta.

Bigi, B. (2015). SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech. *The Phonetician*, 111–112:54–69.

Cathiard, M.-A., Attina, V., and Alloatti, D. (2003). Labial anticipation behavior during speech with and without cued speech. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 1935–1938, Barcelona, Spain.

Clarke, B. R. and Ling, D. (1976). The effects of using cued speech: A follow-up study. *Volta Review*, 78(1):23–34.

Cornett, R. O. (1967). Cued speech. *American annals of the deaf*, pages 3–13.

Duchnowski, P., Braida, L.-D., Bratakos, M.-S., Lum, D.-S., Sexton, M.-G., and Krause, J.-C. (1998). A Speechreading aid based on phonetic ASR. In *5th International Conference on Spoken Language Processing*, pages 3289–3293, Sydney, Australia.

Kaplan, H. (1975). *The effects of cued speech on the speech-reading ability of the deaf.* Ph.D. thesis, ProQuest Information & Learning.

Leybaert, J., Colin, C., and Hage, C., (2010). *Cued speech and cochlear implants*, pages 107–125. 01.

Liu, L., Feng, G., and Beautemps, D. (2018). Automatic temporal segmentation of hand movements for hand positions recognition in french cued speech. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3061–3065, Seoul, South Korea.

Liu, L., Li, J., Feng, G., and Zhang, X.-P. (2019). Automatic Detection of the Temporal Segmentation of Hand Movements in British English Cued Speech. In *Proc. Interspeech*, pages 2285–2289, Graz, Austria.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588):746–748.

Neef, N. A. and Iwata, B. A. (1985). The development of generative lipreading skills in deaf persons using cued speech training. *Analysis and intervention in developmental disabilities*, 5(4):289–305.

Rönnberg, J., Samuelsson, S., Lyxell, B., Campbell, R., Dodd, B., and Burnham, D. (1998). Conceptual constraints in sentence-based lipreading in the hearing-impaired. *The psychology of speechreading and auditory–visual speech*, pages 143–153.

Rönnberg, J. (1995). Perceptual compensation in the deaf and blind: Myth or reality? *Compensating for psychological deficits and declines: Managing losses and promoting gains*, pages 251–274.

Wang, J., Gu, N., Yu, M., Li, X., Fang, Q., and Liu, L. (2021). An attention self-supervised contrastive learning based three-stage model for hand shape feature representation in cued speech. In *Interspeech*, pages 626–630, Brno, Czech Republic.

Weenink, P. B. . D. (1992-2021). Praat: doing phonetics by computer [computer program].

11. Language Resource References

Bigi, B. and Zimmermann, M. (2021). *CLeLfPC - Corpus de Lecture en Langue française Parlée Complétée*. distributed via Ortolang, ISLRN https://hdl.handle.net/11403/clelfpc.

Liu, L. and Hueber, T. and Feng, G. and Beautemps, D. (2018). *French Cued Speech*. distributed via Zenodo, 2.0, ISLRN https://doi.org/10.5281/zenodo.1206001.

Liu, L. (2019). *British English Cued Speech*. distributed via Zenodo, 1.0, ISLRN https://doi.org/10.5281/zenodo.3464212.