User manual for *SyllabO*+

Version August 4th 2016



LABORATOIRE DES NEUROSCIENCES DE LA PAROLE ET DE L'AUDITION

SPEECH AND HEARING NEUROSCIENCE LABORATORY



User manual

Before downloading the full database, it is recommended to have an idea of what data is needed for your project (given its size 177 MB). The databases contain a large amount of, and it is possible that only a portion of it will be useful for a particular purpose.

Instructions

 Go to <u>http://syllabo.speechneurolab.ca/</u> and after filling in a few questions, you will access the *SyllabO*+ database. All the information will remain confidential and will only be used to compute statistics and report to funding agencies.

Accéder à / Go to SyllabO+

2. You may then choose either the complete database (information extracted from all the audio recordings), whether syllables or phones, or a specific database. The complete database contains statistics about the use of phones and syllables across communication contexts, age and gender. It represents the average use of oral language in native adult speakers of Quebec French, in terms of syllables, syllable sequences and phones.

The example below shows the complete syllable database.

Syllable database

When you download the entire syllable database, four files are included in a .ZIP folder: the single syllables, the sequences of two syllables (pairs), sequences of three syllables (triads) and sequences of four syllables (tetrads).

Alternative downloads are also available below in the form of .XLSX files and .CVS files transcribed in SAMPA (that is, without special characters).

Download complete syllable database

Alternatively, you can also download a **specific database**, which will contain statistics computed from a subset of the audio recordings, by using the following parameters, individually or combined: context of communication (formal, informal), age (range), and sex of the speakers. This means that you could obtain the statistics, for example, of syllable use for only the young adults, or for all women in formal contexts, to name a few of the many possibilities. Note that this option is only available for the syllable database at the moment. We expect to provide this option for the phone database during the Fall 2016. Note also that online viewing of the results is currently not available, but we are working on this!

The image below shows the form to download a specific database.

Search in syllable database

Please identify your specific research parameters. The only mandatory field is the type of unit.

Туре*	Syllabes - Syllables
Contexte d'élocution – Communication context	Tous – All
Genre des locuteurs – Speaker gender	Tous – All
Âge minimum – <i>Minimum age</i>	Optionnel – Optional
Âge maximum – Maximum age	Optionnel – Optional
Code de sécurité – Security code*	SZBA
lécharger la base de données spécifique de SYLLABES	- Download specific syllable database

Click to download the desired database.

- 3. If you have **downloaded a complete database**, you will obtain four different files compressed in ZIP format for convenience. For the syllables, there will be one file for the individual syllables and their statistics, one for the pair of syllables and their statistics, one for the groups of 3 syllables (triads) and their statistics, and a last one for the groups of 4 syllables (tetrads) and their statistics. The size of the uncompressed files is 177 MB. For the phones, you will also obtain 4 files, for the individual phones and their statistics, one for the groups of 3 phones (triads) and their statistics, and a last one for the pair of phones and their statistics, one for the groups of 4 phones (tetrads) and their statistics. The size of the uncompressed files is 22 MB.
- 4. If you **downloaded a specific database**, you will only obtain the selected file (syllables, pairs of syllables, or triads).

- 5. All files are in CSV format (Comma Separated Values), which is a simple way of storing tabular data in plain text files. Alternative downloads are also available in the form of .XLSX files (with special IPA characters already embedded) and .CVS files transcribed in SAMPA Speech Assessment Methods Phonetic Aphabet (without any special characters). However, these formats are at the moment only available to download the entire database.
- 6. To open the CSV (API) files we recommend using the software LibreOffice. We recommend it because it handles all phonetic (IPA) characters correctly. Important: Microsoft Excel does not handle these special characters properly. This is why we provide instructions to guide you through the process of having the characters recognized, and this is also why we provide alternative downloads. Note: LibreOffice is a free and open source software available for download at http://www.libreoffice.org/.



7. Select Unicode (UTF-8) as the character set. Select Comma as the separator option. Click OK

to open.

	•		Text Imp	ort - [1_syllabes.cs	v]				
Import	t								
Char	racter set:	et: Unicode (UTF-8)							
Lang	juage:	Default -	English (USA)					
From		1							
FIUI	TTOW.	1							
Separa	ator Optic	ons							
OF	ixed widt	h		💿 Separate	ed by				
	Tab	🔽 Comma	Sen	nicolon 🗌 Spa	ace Other				
	Merge d	elimiters			Text delim	iter:			
Other	Options								
	Quoted fie	ld as text		Detect s	special numbers				
Fields									
Colu	mn type:		0						
	Standard	Standard	Standard	Standard	Standard				
1	Svllahe	Structure	Eréquence	Pourcentage	Rang centile				
2	9	V	8994	2.96209935581	99,9808171878				
3	se	cv	6497	2.13973310148	99,9616343756				
4	de	CV	5245	1.72739727832	99.9424515634				
5 0	də	CV	5156	1.69808586597	99.9232687512				
6	le	CV	4802	1.5814989	99.904085939				
7	е	v	4510	1.48533112016	99.8849031268				
0	10	rv.	1122	1 26116020/81	00 96572031/6	_			
l í									
H	lelp				ОК	Cancel			
A DOMESTIC OF									

- Once the file is opened, you can select only the columns relevant to your project (*Hide* or *Delete* unwanted columns).
- 9. Filter the data according to your needs: select all data, go to the menu, and click on Data -

Filter – Standard Filter.

Ē	• 🖻 •		🔊 🖨 🛛	% D E	• # 15	Select Range Refresh Range		₩ ↓↑	z↓ z↓	SG 💽	. III. II	.
Li	beration S	ians 🔽	10 🖌 📔	B I U I		Streams XML Source	*	<u>↓</u> \$?	% 0.0	00.	.0,	_ ,
A1:/	AMJ10485	576 🔽	f× ∑ =			Cost	-					
	Α	В	С	D	E	Sort		AutoEiltor		1	ī	
1	Syllabe	Structure	Fréquence	Pourcentage	Rang centile	Filler		Standard Fil	tor			F
2	а	V	8994	2.9620993558	99.9808171	Subtotals		Advanced F	ilter	-		Л
3	Se	CV	6497	2.1397331015	99.9616343	Validity		, la fa lo da l				
4	de	CV	5245	1.7273972783	99.942451		-	Reset Filter		-		2
5	də	CV	5156	1.698085866	99.923268	Multiple Operations		Hide AutoFi	ter			
7		V	4602	1 4952211202	99.90408	Text to Columns						Ø
8	la	CV	4510	1 3611602049	99.8657201	Consolidate						
9	la	CV	4107	1 3526064103	99.846537	Group and Outline						Ť×
10	mñ	CV	3763	1 2393128615	99 8273546	Pivot lable						
11	ka	CV	3683	1,2129655245	99,808171	Statistics						
12	ã	V	3598	1.184971479	99,7889890	658						
13	pi	CV	3565	1.1741032025	99.7698062	536						
14	te	CV	3530	1.1625762426	99.7506234	414						
15	pa	CV	3434	1.1309594383	99.7314406	292						
16	õ	V	3285	1.0818875232	99.712257	817						
17	si	CV	3067	1.0100910301	99.6930750	048						
18	sa	CV	2945	0.9699113412	99.6738921	926						1
19	œ	V	2775	0.9139232502	99.6547093	804						
20	sa	CV	2633	0.8671567271	99.6355265	682						
21	mε	CV	2472	0.8141327115	99.616343	756						
22	зе	CV	2431	0.8006297014	99.5971609	438						
23	lə	CV	2406	0.7923961586	99.5779781	316						
24	di	CV	2405	0.7920668168	99.5587953	194						
25	ne	CV	2321	0.7644021131	99.5396125	072						
26	da	CV	2181	0.7182942734	99.520429	695						
21		V	2110	0.6949110119	99.5012468	828						
28	0 kõ	V CV	2088	0.0870054942	99.4820640	700						
30	kj	CV	2022	0.650671//07	99.4020012	162						
31	N	CVC	2003	0.0590714487	00 /2/515	634						
32	siõ	CSV	1913	0.6079648	99 4053328	218						
33	vn	VC	1691	0.5569168346	99.3861500	096						
34	par	CVC	1672	0.5506593421	99.3669671	974						
35	dv	CV	1595	0.5253000303	99.3477843	852						
36	VE	CV	1564	0.5150904372	99.328601	573						
37	tu	CV	1561	0.5141024121	99.3094187	608						
38	me	CV	1556	0.5124557035	99.2902359	486						
39	ma	CV	1555	0.5121263618	99.2710531	364						

10. You may then use any **filter** you wish to obtain the desired information. Select which column is to be filtered, according to which parameter. For example:

		Standard Filter		
Filter Criteria				
Operator	Field name	Condition	Value	
	Structure 🗘	=		
AND ᅌ	Fréquence 🗘	>		
	- none - 🗘 🗘	=	\$ •	
	- none - 🗘 🗘	=	•	
 Options 				
Help		l	OK Cancel	

11. You obtain your **results**!

Example of stimuli creation

Here is an example of how specific stimuli can be created from *SyllabO*+. Using the table containing groups of 3 syllables, these specific columns were selected:

- Group of 3 syllables (IPA transcription)
- Structure 1st syllable
- Structure 2nd syllable
- Structure 3rd syllable
- Percentile of score 1st syllable
- Percentile of score 2nd syllable
- Percentile of score 3rd syllable

The filter " = CV" was applied to all "structure" columns and the filter " >= 95" was applied to all "percentile of score" columns. The resulting table (*Table 1*) is thus a selection of syllable triads, where each syllable has a simple "consonant-vowel" structure and has a high frequency since it is found

between the 95th and 100th percentile. From this table, the groups of 3 syllables can be easily used as stimuli in an experimental study.

Table 1. Stimuli: triads of simple-structure (CV), high-frequency syllables							
Triad	Structure 1	Percentile 1	Structure 2	Percentile 2	Structure 3	Percentile 3	
ko mã se	CV	99.0216765778	CV	99.8273546902	CV	99.9616343756	
vu za ve	CV	98.849031268	CV	98.6188375216	CV	98.273546902	
mã do ne	CV	99.8273546902	CV	97.602148475	CV	99.5396125072	
le zã fã	CV	99.904085939	CV	98.4270093996	CV	97.0650297334	
di fe rã	CV	99.5587953194	CV	96.451179743	CV	98.0625359678	

That's it! Hope you enjoy it, and stay tuned for future developments on the phone database, online searches and the lexical database! Please do not forget to cite our article in all your publications and reports derived from the use of SyllabO+ (Bédard et al., Révisions soumises):

Bédard, P., Audet, A.-M., Drouin, P., Roy, J. P., Rivard, J., & Tremblay, P. (Révisions soumises). SyllabO+: a new tool to investigate sublexical phenomena in Québec spoken French. *Behavioral Research Methods*.