Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 1939–1944
Marseille, 20-25 June 2022

PATATRA and PATAFreq: two French databases for the documentation of within-speaker variability in speech

Cécile Fougeron[*], Nicolas Audibert[*], Cédric Gendrot[*], Estelle Chardenon*[1], Louise Wohmann[*]
[*]Laboratoire de Phonétique et Phonologie (UMR7018 CNRS/Univ. Sorbonne-Nouvelle),
[1]Laboratoire Parole et Langage (UMR 7309, CNRS/AMU Univ.)
[*]19 rue des Bernardins, 75005 Paris, France.
{cecile.fougeron, Nicolas.audibert, cedric.gendrot, louise.wohmann}@sorbonne-nouvelle.fr
estelle.chardenon@univ-amu.fr

Abstract

Our knowledge on speech is historically built on data comparing different speakers or data averaged across speakers. Consequently, little is known on the variability in the speech of a single individual. Experimental studies have shown that speakers adapt to the linguistic and the speaking contexts, and modify their speech according to their emotional or biological condition, etc. However, it is unclear how much speakers vary from one repetition to the next, and how comparable are recordings that are collected days, months or years apart. In this paper, we introduce two French databases which contain recordings of 9 to 11 speakers recorded over 9 to 18 sessions, allowing comparisons of speech tasks with a different delay between the repetitions: 3 repetitions within the same session, 6 to 10 repetitions on different days during a two months period, 5 to 9 repetitions on different years. Speakers are recorded on a large set of speech tasks including read and spontaneous speech as well as speech-like performance tasks. In this paper, we provide detailed descriptions of the two databases and available annotations. We conclude by an illustration on how these data can inform on within-speaker variability of speech.

Keywords: inter- and intra-speaker variation, longitudinal, speech, French

1. Introduction

Variability in speech is pervasive but it is also, at least in part, structured and ruled-governed. Therefore, documenting variation in speech and explaining its origins has been a longstanding endeavor in phonetic research.
Two kinds of variability can be distinguished in terms of their origins and their characteristics: inter-talker and intra-talker variability (Wright 2006). On one side, inter-talker variability is rooted in speaker-specific physiological or anatomical traits, as well as demographical, regional or social factors (Foulkes & Docherty 2006). This inter-talker variability carries indexical information on the speaker: it is disseminated in an array of speech features and it is usually a bundle of speech features which are able to discriminate a speaker (or group of) from another. Furthermore, inter-talker variation is supposed to remain constant across speaking context.
On the other side, intra- or within-taker variability originates from various sources (see for instance Bürki 2018 for a review). Among those, changes in the phonetic realization of phonemes, due to phonetic or prosodic contexts, has been the most documented and this within-talker variability is easy to model. This is not the case of other intra-talker sources of variability which are due to the fact that talkers adapt their speech to the speaking context, to its formality, to the social relation they have with their interlocutors, to the listeners' specific needs (in a noisy context for instance, or whether they share common knowledge on the topic or not), etc. These various factors, often reduced under the notion of 'speech style effects', are mostly impacting the degree of articulatory precision (or hyper/hypo articulation) and the speech rate adopted by the speaker. A third type of factor affecting the way a talker speaks is linked to the speaker affective and cognitive state (e.g. emotion, attention, fatigue…, see Gelfer 1991, Barret & Paus 2002, Johnstone & Scherer 2000).

Although acknowledged and somehow documented, changes in speech resulting from the speaking context or the speaker's state are difficult to delineate and to manipulate experimentally (see for instance the work on 'clear speech' by Scarborough & Zellou, 2013). Moreover, our knowledge of speaker-internal variation is further complicated by the fact that all these effects are not static but rather change dynamically along a discourse (Bates 2003). Indeed, environmental conditions, speech content, prosodic properties affecting pronunciation as well as fatigue, emotions, arousal or attention do vary from one moment to another.
Consequently, it is a common and repeated saying in speech sciences that a token will never be pronounced twice the same way by a single individual. However, little is known on how much speech diverges from one iteration to the next, on which aspects, how fine-grained need the lens be to measure these variation, or what factors contribute to token to token variability (see also Whalen et all 2018). Heald & Nusbaum (2015), for instance, have explored the variability of isolated vowel production for eight speakers over 9 repetitions collected either the same day at three different times or on different days. They found more variability in the vowel acoustics across the different sessions within the same day than across days, and propose a large set of possible accounts for this phenomenon.
In forensic phonetics, where recordings of an individual need to be compared, these questions appear crucial. Indeed, to our knowledge, we have no scientific arguments concerning the validity of a comparison between recordings weeks, months, or years apart and on the conditions allowing such comparisons.
The two databases presented in this paper have been constructed to contribute to the amount of empirical data available to document within-speaker variation in the speech of individuals. Unlike other databases allowing

studies of intra-talker variability on a large set of talkers (among others, Keating et al. 2019: 200 talkers, 12 speech tasks, but only 3 recording sessions), our two databases document variability on a restricted set of speakers (9 to 11 speakers) recorded on various speech tasks ('laboratory' speech material, speech-like tasks and in spontaneous speech), with a focus on intra-speaker variability over a longer time lapse: speakers are recorded over multiple sessions (7 to 18 sessions according to the speakers) and with different delays between recordings.

2. Description of the PATATRA database

2.1. Speakers and database content

The collection of the PATATRA (Parole AdulTe A TRavers les Ages – *adult speech across ages*) database has started in 2013 at the L.P.P. laboratory. The purpose of this long-term project was to constitute a longitudinal database of recordings on a selected set of speakers, all non-naïve since they are all working on speech, but easily available for multiple recordings over the years.
To date, the database includes a total of 92 recording sessions, produced by 11 speakers, all researchers in speech sciences (7 female & 4 male speakers), which are recorded almost every year. As shown in Table 1, speakers are of different ages and cumulates different numbers of recording sessions (5 to 9 sessions), because they have not been included in the project at the same time or have missed some recordings. Most of the speakers are native French speakers, with a Parisian/North of France accent, and two of them (annotated with a * in Table 1) are fluent but not French-native.
For almost all the recording sessions (except for the year 2016 and 2020), the available data comprise the audio signal and a synchronously recorded EGG (electroglottographic) signal.
Recordings are completed by a self-assessment of the speaker voice quality on the day of recording with a French version of the French version of the Voice Handicap Index questionnaire (Woisard et al. 2004) and information about potential smoking and drinking habits and ear/nose/throat infections during the year are provided.

SPK	sex	N	year1	missed	age @ year1	Cumul sessions
F01	F	9	2013	N/A	39	16
F02	F	8	2013	2015	42	17
F03	F	9	2013	N/A	42	18
F04	F	8	2014	N/A	68	13
F05	F	9	2013	N/A	57	16
F06*	F	9	2013	N/A	52	9
F07	F	5	2017	N/A	36	14
H01	M	9	2013	N/A	36	16
H02	M	9	2013	N/A	35	18
H03	M	9	2013	N/A	62	18
H04*	M	8	2013	2016	42	8

Table 1: Available data in the PATATRA dataset. *= near French native speakers; N=number of available sessions, year1=first year of recording, missed=missing years, age@year1=age of the speaker at the date of the first recording. The last column indicates the total number of recorded sessions for each speaker cumulating both the PATATRA and the PATAFreq protocols.

	V=/i, a, u/		V=/i, a, u/
/b_k/	bic, bac, bouc	/t_ʁ/	tire, tard, tour
/b_z/	bise, base, bouse	/t_f/	tiffe, taffe, touffe
/b_l/	bile, balle, boule	/d_t/	dite, date, doute
/b_ʃ/	biche, bache, bouche	/l_ʁ/	lire, lard, lourd
/b_f/	biffe, baffe, bouffe	/l_v/	live, lave, louve
/p_l/	pile, pâle, poule	/ʒ_t/	gite, jatte, joute
/p_ʁ/	pire, par, pour	/k_ʁ/	kir, car, cours
/m_l/	mille, malle, moule	/k_ʃ/	quiche, cache, couche
/m_ʃ/	miche, mâche, mouche	/ʁ_t/	rite, ratte, route

Table 2: List of the CVC 18 minimal triplets in the PATATRA protocol, with V: /i, a, u/

2.2. Recording procedure: audio and EGG signals

Each year, recordings of the PATATRA protocol are done in the LPP sound booth by a research assistant. Speakers are equipped with a AKG C520 head mounted microphone and an electroglottograph device (Glottal Entreprise, EG2-PCX2), with the collar placed according to the position of the speaker's vertebrae after applying conductive gel on the electrodes.
The audio and EGG signals are captured on the two channels of a Digidesign Digi003 Rack soundcard piloted with the Protool software.
Due to COVID-19 restrictions, the 2020 recording could not take place in the same conditions. We have thus considered the first recording of the PATAFreq database, which includes part of the same speech material (see section 3), as the recording of year 2020.

2.3. Speech material

The recording sessions always follow the same protocol, with the five speech and speech-like tasks presented in the order given below. The content of the tasks is presented in the Appendix.

- [Txt3] The speaker reads a short text (the French version of the North wind and the sun, 'La Bise et le Soleil') three times successively. S/he is instructed to get familiarized with the text before starting, to read aloud at comfortable pace, intensity and pitch, and to repeat the full sentence in case of disfluencies.
- [Lst1] The speaker reads a list of 56 French monosyllabic words three times successively. Except for the first and the last words (introduced to anchor the starting and ending intonation), the words form 18 minimal triplets allowing the occurrence of the vowels /i, a, u/ in the same

consonantal context, as shown in Table 2. The speaker is instructed to read the words one by one, at comfortable pace intensity and pitch, with a silent pause between consecutive words. This corpus is meant to capture the i/a/u vowel space and varied coarticulatory effects between vowels and consonants.

- [Spont] The speaker produces around five minutes of spontaneous speech with the research assistant, whose role is to entertain the conversation by asking questions on free topics (holidays, movies, work, etc…).
- [Gliss] Then, two f0 glissandi at low and high intensity are collected in order to capture aspects of the dynamic range of the voice. The speaker is instructed to produce a siren from the lowest frequency to the highest on a /a/ vowel, first at the lowest possible intensity, then the highest. A model can be provided by the research assistant if the instruction is not understood.
- [MPT] Finally, a task meant to capture the maximum phonation time of the speaker is performed. The speaker is instructed to produce, after a deep breath, a /a/ sound as long as possible, at a comfortable pitch and intensity. Three trials are recorded in succession, and the longest one is kept as a measure of pneumo-phonatory control.

In this protocol, within-speaker longitudinal comparisons can be assessed by comparing all the tasks across the years of recordings. Furthermore, within-speaker variation intra-session can be assessed by comparing the three repetitions of the tasks [text3] and [Lst1].

3. Description of the PATAFreq database

2.1. Speakers and database content

The PATAFreq database was created during the first lock-down in 2020 by 8 of the PATATRA speakers who recorded themselves on a regular basis, with a goal of about 10 recordings over a two months period.

The 9 speakers are the native French speakers of the PATATRA database and are indexed with the same code in Table 1 and 2.

As shown in Table 3, the database contains a total of 80 recordings, with 8 to 10 sessions per speakers, except for speaker F04 who encountered technical problems and could do only 6 sessions. Speakers were instructed to record themselves on a regular basis, with a minimum of 24h between sessions and in no more than a two months period. All recording sessions are indexed with the date and time of recording. As shown in Table 3, the average lag between successive recordings vary slightly between the speakers but all sessions have been recorded done within a month and a half. Speakers were encouraged to change the time of the day for the recordings but family constraints during the lockdown did not always permitted this.

Recordings are completed by a self-assessment on a four-point scale of the participant’s overall fatigue, emotional state, vocal fatigue and amount of voice use on the day of recording.

2.2. Recording procedure: audio signal only

All speakers were equipped with professional sound cards and microphones adapted to their different computer configurations. The recording material used is indicated in Table 3 by speaker. They were instructed to record themselves in a quiet environment, always in the same room if possible, and with the same settings. Any change, or unexpected environmental noise, had to be signaled in a questionnaire to be filled after each recording. At the beginning and at the end of each recording session, 5 seconds of silence is recorded in order to control for a change in ambient noise if needed.

Most speakers managed their recordings with the help of a dedicated app developed in Python (a modified version of MonPaGe, Trouville et al. 2020) which took care of prompting the instructions, recording and organizing the sound files. Speaker F04 and F03 could not use this app and did their recordings with Praat, relying on a Powerpoint version of the prompts.

2.3. Speech Material

The eight speakers of the PATAFreq database also participate to the PATATRA project and the idea was to reproduce a similar (but not identical) protocol which was convenient and short enough (10-15 minutes) to be recorded on oneself at home several times a week.

For each session, seven speech and speech-like tasks were recorded in the following order:

- [Txt1,2,3]: The speaker reads once, at a comfortable pace, pitch and intensity, three short texts (text1, text2, text3, see appendix) in a row. Txt3 is ‘La Bise et le Soleil’ which is also included in the PATATRA protocol. Speakers are instructed to get familiarized with the text before reading aloud and to repeat the full sentence in case of disfluencies.
- [Lst2]: The speaker reads a list of 26 French monosyllabic words and non-words constructed around the 13 French vowels (/i, e, ɛ, a, y, ø, œ, u, o, ɔ, ɑ̃, ɛ̃, ɔ̃/) in a /t_(t)/ and /k_(k)/ contexts (see appendix for the full list). The items are presented in an orthographic and IPA form in order to assure the production of the targeted vowel. This list is aimed at documenting variability in the production of a large set of French vowels (oral and nasal) and in the coarticulatory effects of back vs. front consonantal contexts.
- [Sent] The speaker reads once a fully voiced sentence (‘Mélanie vend du lilas’ – *Melanie sells lilac*), on which speaking f0 and articulation rate can be measured.
- [Spont] Since the speakers were recording themselves with no interlocutor, we designed a procedure allowing to capture some continuous, semi-spontaneous speech, in response to two questions prompted on the screen. The first one, with a narrow topic is ‘How do you make an omelet?’, for which the speaker is encouraged to give a lot of details and also to vary recipes along the sessions. The second one has a larger topic since the speaker is invited to tell what s/he has been doing since the previous recording. Dealing with non-naïve speakers here was crucial since

they knew why they had to fill-up with speech the 2-3 minutes timer presented on the screen for each question.

- [Sust] The speaker is instructed to produce a sustained /a/ for 2-3 seconds at comfortable pitch and intensity. This sustained vowel is meant to be used to measure voice quality parameters.
- [Gliss] The speaker produces the glissando at low intensity of the PATATRA protocol.
- [MPT] Then s/he produces two trials of the maximum phonation time of the PATATRA protocol.
- [DDK] Finally, the speaker is asked to produce as fast and as clearly as possible, a succession of syllables. This diadochokinetic task (used in clinical settings) is a maximal performance task testing the ability to produce alternating speech movements under time and precision constraints. The items to be produced include the succession of identical syllables: /bababa…/ then /gogogo…/; and then a alternation of different syllables /badego…/.

The tasks [Txt3], [Spont], [Gliss], [MPT] are directly comparable to the ones recorded for the PATATRA protocol. The other tasks have been introduced to test for within-speaker variability on aspects of speech/speech-like performances ([DDK], [Sent], [Sust]) used in clinical protocols and for which we have comparable references on more than 400 French speakers within the MonPaGe_HA database (Fougeron et al. 2018).

SPK	Sex	N	Age	Lag	Sound Card	Syst	Mic
F01	F	8	46	3.6	Presonus iTwo	OS 10.15	AKG C420L
F02	F	10	50	4.1	Roland UA-22	OS 10.13	AKG C520L
F03	F	10	50	3	Edirol UA-25	OS 10.6	AKG C520
F04	F	6	77	9.4	Edirol UA-25	OS 10.13	AKG C420L
F05	F	8	64	2.3	Roland UA-22	OS 10.13	AKG C520L
F07	F	10	40	3.5	Presonus iTwo	OS 10.12	AKG C520L
H01	M	8	44	6.3	Edirol UA-25	Windows 10 Pro	AKG C520
H02	M	10	43	3.6	Focurisite Scarlett 2i4	OS 10.14	AKG C520
H03	M	10	70	3.2	Presonus iTwo	OS 10.11	AKG C420L

Table 3: Available data in the PATAFreq dataset and recording material (sound card, system and microphone) used. Lag= mean lag between recordings (in days).

4. Available Annotations

To date, the read part of the two databases are annotated with manual segmentation of the texts into defined chunks and inter pauses units (IPU). In order to allow paradigmatic comparisons across recordings on the same part of speech, but also syntagmatic comparisons between part of speech with equivalent number of syllables along the text reading, several chunks have been pre-defined on the three text. These 10 to 18 chunks, according to the text, are presented in Table 4, and comprise 8 to 15 syllables.

Within these chunks, pauses and interpausal units have been manually segmented, with the following conventions: when pauses are located at chunk boundaries, they are included in the previous chunk; when disfluencies are present in the reading, if the speaker repeat the sentence the disfluent part is discarded in favor of the repeated one, otherwise they are included and transcribed within a chunk; all perceived pauses are segmented as pauses whatever their length.

Alignment of the interpausal units has been done with WebMAUS (Schiel, 1999; Kisler et al., 2017) but not manually corrected yet.

Automatic transcription and manual verification of the spontaneous part of the databases is currently under process.

	#	SAMPA	ph	syl
text1	1	o nɔʁ dy pɛi ɔ̃ tʁuv yn espɛs də	23	11
	2	ʃa dɔ̃ la kø ɛ tʁɛ kuʁt il	18	8
	3	sɔ̃ nwaʁ avɛk dø taʃ blɑ̃ʃ syʁ lə do	26	10
	4	lœʁ pwal ɛ bo e du ʒyst a kote	22	10
	5	vi yn koloni dwazo dɔ̃ le ni	21	10
	6	sɔ̃t akʁoʃe o bɔʁ də la falɛz	22	10
	7	il dwav fɛʁ atɑ̃sjɔ̃ a nə pa fɛʁ tɔ̃be lœʁ zø	32	12
	8	dɑ̃ la mɛʁ ma sœʁ na ka tʁavɛʁse	24	12
	9	la ʁy puʁ ʁɑ̃kɔ̃tʁe se dø zespɛs	24	10
	10	vivɑ̃ ɑ̃ naʁmoni o kœʁ dɛ̃ paʁk natyʁɛl	29	13
	11	ʁegyljɛʁmɑ̃ syʁ lə ku də midi	23	10
	12	apʁɛ zavwaʁ pʁi ɛ̃ bɔ̃ te nu sɔʁtɔ̃ də	27	12
	13	ʃe zɛl puʁ ale obsɛʁve se zanimo	26	12
text2	1	ma sœʁ ɛ vəny ʃe mwa ijɛʁ puʁ pʁɑ̃dʁə lə te	32	13
	2	ɛl mə paʁlɛ də se vakɑ̃s ɑ̃ mɛʁ dy nɔʁ	27	12
	3	lɔʁskə dɑ̃ notʁ do tɔ̃ba ɛ̃ pəti twazo	28	13
	4	se dø zɛl etɛ blese e il	18	9
	5	avɛ ʁəsy ɛ̃ ku vjolɑ̃ syʁ la kø	22	11
	6	sɔ˜kœʁ batɛ tʁɛ vit mɛ il etɛ tɑ̃ vi	26	12
	7	sɔ̃ plymaʒ etɛ bo e du ʒə mapʁoʃɛ	25	13
	8	dy bɔʁ də la fənɛtʁ puʁ ʁəgaʁde dɑ̃ la ʁy	31	13
	9	ɛ̃ ʃa selwaɲɛ dɛ̃ ni pɛʁʃe syʁ ɛ̃ naʁbʁ	28	12
	10	il avɛ dy fɛʁ fɥiʁ lwazo apʁɛ lavwaʁ atake	34	15
text3	1	la biz e lə solɛj sə dispytɛ	22	10
	2	ʃakɛ̃ asyʁɑ̃ kil etɛ lə ply fɔʁ	23	11
	3	kɑ̃t ilz ɔ̃ vy ɛ̃ vwajaʒœʁ ki	19	9
	4	savɑ̃sɛ ɑ̃vəlope dɑ̃ sɔ̃ mɑ̃to	21	10
	5	il sɔ̃ tɔ̃be dakɔʁ kə səlɥi	20	9
	6	ki aʁivəʁɛ lə pʁømje a	18	9
	7	fɛʁ ote sɔ̃ mɑ̃to o vwajaʒœʁ	21	10
	8	səʁɛ ʁəgaʁde kɔm lə ply fɔʁ	22	9
	9	alɔʁ la biz sɛ miz a sufle	20	9
	10	də tut se fɔʁs mɛ plyz ɛl suflɛ	24	9
	11	ply lə vwajaʒœʁ seʁɛ sɔ̃ mɑ̃to	23	10
	12	otuʁ də lɥi e a la fɛ̃	15	8
	13	la biz a ʁənɔ̃se a lə lɥi fɛʁ ote	24	12
	14	alɔʁ lə solɛj a komɑ̃se a bʁije	24	12

	15	e o bu dɛ̃ momɑ̃ lə vwajaʒœʁ	20	10
	16	ʁeʃofe a ote sɔ̃ mɑ̃to	16	9
	17	ɛ̃si la biz a dy ʁəkonɛtʁ	19	9
	18	kə lə solɛj etɛ lə ply fɔʁ de dø	24	11

Table 4: Definition of the chunks for the 3 text, with API transcriptions and number of phonemes and syllable per chunk.

5. First explorations and conclusion

Explorations of the databases are currently under process in the context of several projects studying within- vs. between-speaker variability in speech and the long-term changes in adult speech with age.

In a first set of studies on the PATATRA database (Chardenon et al., 2020, Audibert et al., 2021), we explored the productions of the read task [Txt3] for 8 of the speakers (F1-5, H1-3) over 7 years. Various speech dimensions have been documented: speech and articulation rate, voiced ratio (cumulated duration of voiced segments), speaking f0, speaking f0 range, LTAS slope. These dimensions have been compared across recordings both as a descriptor of the whole recording (mean value for each dimension), but also as a descriptor of the modulation within a single recording for the specific dimension (variance across successive chunks).

Results show that intra-speaker variability on temporal dimensions was found to be more important between distant recordings than between successive repetitions within the same year session. Within speaker variability appears to be speaker-specific. Some speakers are much more variable from one recording to the next on certain dimensions (e.g., rate of speech, or pitch) while others are more stable. Of particular interest, speakers also vary on how the modulate their speech within a single recording. Over these 7 years of recordings, no specific trend that could be related to aging has been observed.

The PATATRA and PATAFreq databases offer a substantial amount of original data documenting intra-speaker variability in speech for a selected set of French speakers. It allows a comparison between recordings of the same ‘laboratory’ speech material across repetitions within the same session, across sessions collected on different days during a two months period and across sessions collected on successive years (9 at the date of this publication). Spontaneous speech for the same speakers can also be compared on repetitions produced years or days apart.

Acknowledgements

The construction of these databases was supported by the program "Investissements d’Avenir" ANR-10-LABX-0083 (Labex EFL) and the VoxCrim project of the French National Research Agency (ANR-17-CE39-0016).

Bibliographical References

Audibert, N., Fougeron, C., & Chardenon, E. (2021). Do you remain the same speaker over 21 recordings? XVII AISV, Zürich.

Barrett J., & Paus T. (2002). Affect-induced changes in speech production. *Experimental Brain Research*, 146(4), 531–537. PMID: 12355282

Bürki A. (2018) Variation in the speech signal as a window into the cognitive architecture of language production. *Psychonomic Bulletin & Review* 25:1973–2004 https://doi.org/10.3758/s13423-017-1423-4

Chardenon, E., Fougeron, C., Audibert, N., Gendrot, C. (2020). Dis-moi comment tu varies ton débit, je te dirai qui tu es. 31e Journées d'Études sur la Parole. Nancy, France. pp.82-90.

Fougeron, C., Delvaux, V., Menard, L., & Laganaro, M. (2018). The MonPaGe_HA database for the documentation of spoken French throughout adulthood. *Proceedings of LREC 2018*, Miyazaki, Japan, pp. 4301-4306.

Gelfer M.P. (1991). Effects of prolonged loud reading on selected measures of vocal function in trained and untrained singers. *Journal of Voice*, 5, 15–167.

Heald SLM, Nusbaum HC (2015) Variability in Vowel Production within and between Days. *PLoS ONE* 10(9): e0136791. doi:10.1371/journal. pone.0136791

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. *Handbook of emotions*, 2, 220-235.

Keating, P. A., Kreiman, J., and Alwan, A. (2019). “A new speech database for within- and between-speaker variability,” in Proceedings of the ICPhS XIX.

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347.

Scarborough, R., & Zellou, G. (2013). Clarity in communication:“Clear” speech authenticity and lexical neighborhood density effects in speech production and perception. *JASA*, 134(5), 3793-3807

Schiel, F. (1999). Automatic Phonetic Transcription of Non-Prompted Speech. Proc. ICPhS, pp. 607-610.

Trouville, R., Delvaux, V., Fougeron, C., Laganaro, M., (2021). Logiciel d’évaluation de la parole (version screening) MonPaGe-2.0.s [Computer program] Retrieved on 05/01/2022 from https://lpp.in2p3.fr/monpage/

Woisard, V., Bodin, S., & Puech, M. (2004). The Voice Handicap Index: impact of the translation in French on the validation. Revue de laryngologie-otologie-rhinologie, 125(5), 307-312.

Whalen, D. H., Chen, W. R., Tiede, M. K., and Nam, H. (2018). Variability of articulator positions and formants across nine English vowels. *J. Phonetics* 68,1–14. doi: 10.1016/j.wocn.2018.01.003

Appendix

1: [Txt3] La bise et le soleil. Sole text of the PATATRA protocol and 3rd text of the PATAFreq protocol.

La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort, quand ils ont vu un voyageur qui s'avançait, enveloppé dans son manteau.

Ils sont tombés d'accord que celui qui arriverait le premier à faire ôter son manteau au voyageur serait regardé comme le plus fort.

Alors, la bise s'est mise à souffler de toute sa force mais plus elle soufflait, plus le voyageur serrait son manteau autour de lui et à la fin, la bise a renoncé à le lui faire ôter.

Alors le soleil a commencé à briller et au bout d'un moment, le voyageur, réchauffé, a ôté son manteau.

Ainsi, la bise a du reconnaître que le soleil était le plus fort des deux.

2. [Lst1]Monosyllabic word list of the PATATRA protocol

Tic, bouc, biffe, pile, poule, baffe, louve, tard, tire, boule, couche, route, bouche, tour, taffe, tiffe, touffe, datte, moule, lard, dite, par, doute, jatte, gîte, car, kir, cours, cache, malle, quiche, bouffe, base, bise, bouse, bic, balle, bile, mâche, bâche, lire, biche, miche, joute, mouche, pire, rite, pour, pâle, bac, ratte, mille, lourd, lave, live, bar

3: [Txt1,2] Text 1 and 2 of the PATAFreq protocol.

Texte 1:

Au nord du pays, on trouve une espèce de chats dont la queue est très courte. Ils sont noirs avec deux tâches blanches sur le dos. Leur poil est beau et doux. Juste à côté, vit une colonie d'oiseaux dont les nids sont accrochés au bord de la falaise. Ils doivent faire attention à ne pas faire tomber leurs oeufs dans la mer. Ma sœur n'a qu'a traverser la rue pour rencontrer ces deux espèces vivant en harmonie au cœur d'un parc naturel. Régulièrement, sur le coup de midi, après avoir pris un bon thé, nous sortons de chez elle pour aller observer ces animaux.

Texte 2:

Ma sœur est venue chez moi hier pour prendre le thé. Elle me parlait de ses vacances en mer du nord, lorsque, dans notre dos, tomba un petit oiseau. Ses deux ailes étaient blessées, et il avait reçu un coup violent sur la queue; son cœur battait très vite mais il était en vie. Son plumage était beau et doux. Je m'approchais du bord de la fenêtre pour regarder dans la rue. Un chat s'éloignait d'un nid perché sur un arbre. Il avait dû faire fuir l'oiseau après l'avoir attaqué.

[Lst2] Monosyllabic word and non-word list of the PATAFreq protocol

taux	[to]	coq	[kɔk]
queue	[kø]	conque	[kɔ̃k]
canque	[kɑ̃k]	tœut	[tœt]
tute	[tyt]	tante	[tɑ̃t]
teinte	[tɛ̃t]	qué	[ke]
kouk	[kuk]	toute	[tut]
tite	[tit]	cuk	[kyk]
quik	[kik]	coke	[kok]
tate	[tat]	kink	[kɛ̃k]
cac	[kak]	kœuk	[kœk]
tote	[tɔt]	teu	[tø]
tête	[tɛt]	tonte	[tɔ̃t]
quêk	[kɛk]	thé	[te]