



Activity Report 2023

Team EXPRESSION

Expressiveness in Human Centered Data/Media

D6 – Media and Interaction Department



1 Team composition

1.1 Composition

Head of the team

Pierre-François Marteau, Professor, Université Bretagne Sud

Administrative assistants

Anne Idier, Université Bretagne Sud

Martine Milcent, Université Bretagne Sud

Angélique Le Pennec, Université de Rennes 1

Joëlle Thépault, Université de Rennes 1

Permanent members

← Pierre Alain, Research Engineer, Université de Rennes 1

Nelly Barbot, Associate professor, Université de Rennes 1

Vincent Barraud, Associate professor, Université de Rennes 1

Nicolas Béchet, Associate professor, Université Bretagne Sud

Giuseppe Bério, Professor, Université Bretagne Sud

Jonathan Chevelu, Associate professor, Université de Rennes 1

Arnaud Delhay, Associate professor, Université de Rennes 1

Sylvie Gibet, Professor, Université Bretagne Sud

Caroline Larboulette, Associate professor, Université Bretagne Sud

Damien Lolive, Professor, Université de Rennes 1

Gildas Ménier, Associate professor, Université Bretagne Sud

→ Vincent Segonne, Associate professor, Université Bretagne Sud

Jeanne Villaneau, Associate professor (emeritus), Université Bretagne Sud

External collaborator

Elisabeth Delais-Roussarie, Senior researcher, CNRS/LLF

Farida Said, Associate professor, Université Bretagne Sud

Non-permanent members

← Gaëlle Vidal, Engineer, Université de Rennes 1 (from April 2019 to March 2023)

← Rashedur Rahman, Post-doctoral researcher, Université de Rennes 1 (from May 2020 to May 2023)

⇒ Aghilas Sini, Post-doctoral researcher, University of Rennes 1 (until January 2022 and from October 2022 until August 2023)

← Antoine Perquin, Post-doctoral researcher, University of Rennes 1 (from December 2021 to November 2023)

⇒ Clément Reverdy, ATER, Université de Bretagne Sud (ATER from November 2022 to August 2023, then postdoc since September 2023)

David Guennec, contractual assistant professor, Université de Rennes (since September 2022)

→ Philippe Martin, Research Engineer, University of Rennes 1 (since June 2023)

PhD students

- ← Alexis Blandin, Université Bretagne Sud, CIFRE UNEEK, defended the 27th of March 2023.
Mansour Tchénégnon, Université Bretagne Sud, CDE/ARED 3rd year
- ← Somayeh Jafaritazehjani, Université de Rennes 1, CDE+TU Dublin, defended in December 2023
- Clémence Mertz, Université Rennes 1, CDE+LTC, 3rd year, gap year
Lily Wadoux, Université de Rennes 1, CDE, 3rd year
- ← Olivier Zhang, Université de Rennes 1, CIFRE Orange Labs, 3rd year, defended the 21st of December 2023
Danrun Cao, Université Bretagne Sud, CIFRE OctopusMind 2nd year
Hoan My TRAN, Université de Rennes 1, CD22+CDE-IA, since November 2022
- Ulysse Oliveri, University of Rennes, CIFRE Airbus, since March 2023
- Pauline Mas, University of Rennes, CIFRE Voxygen, since September 2023
- Quentin Lemesle, University of Rennes, CDE+DGA, since October 2023
- Rania Bennetayeb, Université Bretagne Sud, CIFRE Rennes Metropole, since December 2023

1.2 Evolution of the staff

The permanent staff increases slightly with the arrival of Vincent Segonne, assistant professor in Vannes. The number of PhD students is stable.

2 Overall objectives

Expressivity or expressiveness are terms which are often used in a number of domains. In biology, they relate to genetics and phenotypes, whereas in computer science, expressivity of programming languages refers to the ability to formalize a wide range of concepts. When it comes to human expressivity, we will consider the following reading: expressivity is the way a human being conveys emotion, style or intention. Considering this definition, the EXPRESSION team focuses on studying human language data conveyed by different media: gesture, speech and text. Such data exhibit an intrinsic complexity characterized by the intrication of multidimensional and sequential features. Furthermore, these features may not belong to the same representation levels - basically, some features may be symbolic (e.g., words, phonemes, etc.) whereas others are digital (e.g., positions, angles, sound samples) - and sequentiality may result from temporality (e.g., signals).

Within this complexity, human language data embed latent structural patterns on which meaning is constructed and from which expressiveness and communication arise. Apprehending this expressiveness, and more generally variability, in multidimensional time series, sequential data and linguistic structures is the main proposed agenda of EXPRESSION. This main purpose comes to study problems for representing and characterizing heterogeneity, variability and expressivity, especially for pattern identification and categorization.

The research project targets the exploration and (re)characterization of data processing models in three mediated contexts:

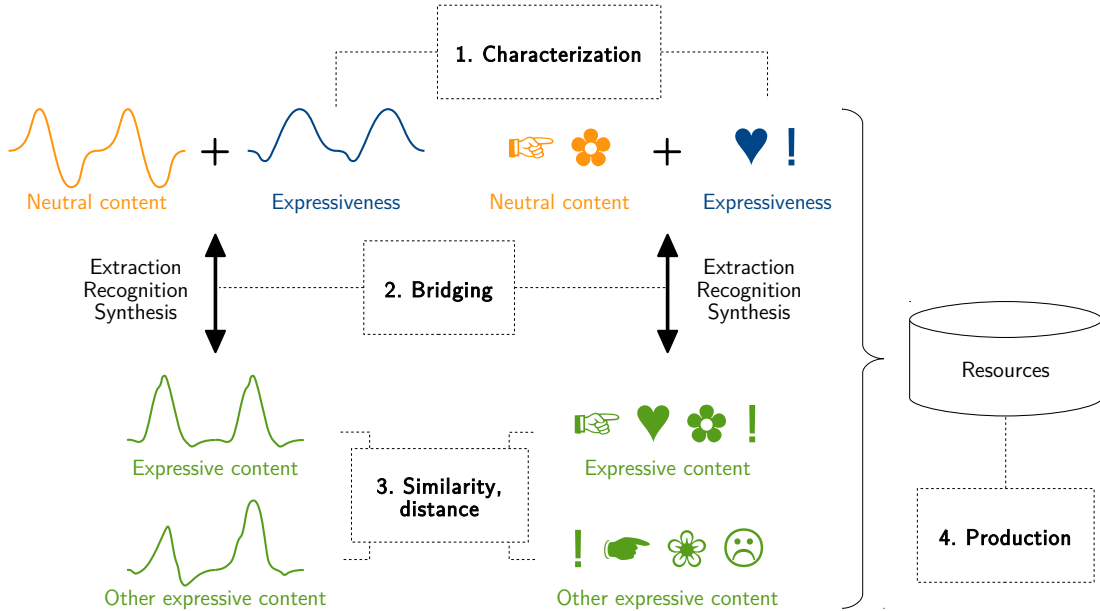


Figure 1: Overview of the main challenges considering both on continuous numerical (left) and discrete symbolic (right) data.

1. Expressive gesture analysis, synthesis and recognition,
2. Expressive speech analysis and synthesis,
3. Expression in text and language.

2.1 Main challenges addressed by the team

Four main challenges will be addressed by the team.

C1: The characterization of the expressiveness as defined above in human produced data (gesture, speech, text) is the first of our challenges. This characterization is challenging jointly the extraction, generation, or recognition processes. The aim is to develop models for manipulating or controlling expressiveness inside human or synthetic data utterances.

C2: Our second challenge aims at studying to what extent innovative methods, tools and results obtained for a given media or for a given pair of modality can be adapted and made cross-domain. More precisely, building comprehensive bridges between discrete/symbolic levels (meta data, semantic, syntactic, annotations) and mostly continuous levels (physical signals) evolving with time is greatly stimulating and nearly not explored in the different scientific communities.

C3: The third challenge is to address the characterization and exploitation of data-driven embeddings¹ (metric or similarity space embeddings) in order to ease post-

¹Given two metric or similarity spaces (X, d) and (X', d') , a map $f : (X, d) \rightarrow (X', d')$ is called an embedding.

processing of data, in particular to reduce the algorithmic complexity and meet the real-time or big-data challenges. The characterization of similarity in such embeddings is a key issue as well as the indexing, retrieval, or extraction of sub-sets of data relevant to user's defined tasks and needs, in particular the characterization of expressiveness and variability.

C4: The fourth challenge is to contribute to the production of resources that are required, in particular to develop, train and evaluate machine learning (statistical or rule-based) models for human language data processing. These resources are mainly corpora (built from speech, text and gesture time series), dictionaries, and semantic structures such as ontologies.

All the addressed challenges are tackled through the development of models, methods, resources and software tools dedicated to represent and manage gesture, speech or textual data. Thus we consider a complete processing chain that includes the creation of resources (corpus, thesaurus, semantic network, ontology, etc.), the labeling, indexing and retrieval, analysis and characterization of phenomena via classification and extraction of patterns (mostly sequential).

These challenges also target multi-level aspects, from digital tokens to semantic patterns, taking into account the complexity, the heterogeneity, the multi-dimensionality, the volume, and the nature of our temporal or sequential data.

We are aiming at addressing these challenges in terms of development and exploitation of machine learning and pattern discovery methods for clustering, classification, interactive control, recognition, and production of content (speech signals, texts or gestures), based on different levels of representation (captured or collected data but also knowledge that is specific to the media or the considered application). Finally, both objective and subjective (perceptive) evaluations of these models are a key issue of the research directions taken by the EXPRESSION team.

2.2 Main research focus

Five thematic lines of research are identified to carry out this research.

RF1: Data acquisition – Gesture, speech or text data are characterized by high levels of heterogeneity and variability. Studying such media requires high quality data sets appropriate to a well defined and dedicated task. The data acquisition process is thus a crucial step since it will condition the outcomes of the team research, from the characterization of the studied phenomena, to the quality of the data driven models that will be extracted and to the assessment of the developed applications. The production of high quality and focused corpora is thus a main issue for our research communities. This research focus addresses mainly the fourth challenge;

RF2: Multi-level representations – We rely on multi-level representations (semantic, phonological, phonetic, signal processing) to organize and apprehend data. The heterogeneity of these representations (from metadata to raw data) prevents us from using standard modeling techniques that rely on homogeneous features.

Building new multi-level representations is thus a main research direction. Such representations will provide efficient information access, support for database enrichment through bootstrapping and automatic annotation. This research focus contributes mainly to the second, third and fourth challenges;

RF3: Knowledge extraction – This research addresses data processing (indexing, filtering, retrieving, clustering, classification, recognition) through the development of distances or similarity measures, rule-based or pattern-based models, and machine learning methods. The developed methods will tackle symbolic data levels (semantic, lexical, etc.) or time series data levels (extraction of segmental units or patterns from dedicated databases). This research focus contributes mainly to the first and third challenges.

RF4: Generation – We are also interested in the automatic generation of high-quality content reproducing human behavior on two modalities (gesture and speech). In particular, to guarantee adequate expressiveness, the variability of the output has to be finely controlled. For gesture, statements and actions can be generated from structural models (composition of gestures in French sign language (LSF) from parameterized linguistic units). For speech, classical approaches are data-driven and rely either on speech segment extraction and combination, or on the use of statistical generation models. In both cases, the methods are based at the same time on data-driven approaches and on cognitive and machine learning control processes (e.g., neuromimetic). This research focus contributes mainly to the first and fourth challenges since generation can be seen also as a bootstrapping method. As parallels can be possibly drawn between expressive speech and expressive movement synthesis, the focus also contributes to the second challenge;

RF5: Use cases and evaluation – The objective is to develop intuitive tools and in particular sketch-based interfaces to improve or facilitate data access (using different modes of indexing, access content, development of specific metrics, and graphical interfaces), and to integrate our aforementioned models into these tools. As such, this focus contributes to the first challenge and has a direct impact on the fourth challenge. Furthermore, whereas many encountered sub-problems are machine learning tasks that can be automatically evaluated, synthesizing human-like data requires final perceptive (*i.e.* human) evaluations. Such evaluations are costly and developing automatic methodologies to simulate them is a major challenge. In particular, one axis of research directly concerns the development of cross-disciplinary evaluation methodologies. This research focus contributes also to the second challenge;

3 Scientific achievements

3.1 New Results by Key Issues

In accordance with the Team Project, the main outcomes for 2023 are listed into the following key issues items defined above for the team:

Use cases
and eval-
uation,
Knowledge
extraction

Building of a gender-annotated corpus using zero-shot learning. [9] In order to best adapt to new technologies, an association has developed a webchat application allowing anyone to express and share their anxieties. Several thousand anonymous conversations have then been brought together and form an unprecedented corpus of stories about human distress and social violence. We present in this paper a methodology to produce a learning model that allows an automatic gender labeling of a corpus of texts in French. The method is based on a combination of a Zero-Shot classification algorithm, human validation, and supervised learning. This method allows us to effectively pre-annotate a large corpus by presenting some experimental results so that an expert can finally more easily validate the annotation produced.

Use cases
and eval-
uation,
Knowledge
extraction

Comparative study of lexical embeddings for the extraction of named entities in French. [10] In this paper we present a comparative study of word embedding methods for French named entity recognition (NER). Our goal is to compare each method's performance when facing the same task and under the same working conditions. We use the French proportion of WikiNER as corpus of study. It is a 3.5-million token corpus with 4 entity types. 10 embedding methods are studied, including non-contextual ones, contextual ones and transformer-based ones. For each embedding, we train a BiLSTM-CRF as token classifier. For transformer-based models, we also compare their performance under another usage which is fine-tuning.

Use cases
and eval-
uation,
Knowledge
extraction

Place, role (s) and form (s) of reformulation in the relationship with a web chat. [7] It is through a corpus of conversations in the form of webchat intended to welcome individuals in distress that we approach the notion of reformulation because callers who reconnect on the platform (we will call them poly-callers) are brought to reconfigure their statements due to the alternation of voluntary listening; their return to the platform means the non-improvement of the caller's condition and the reformulation helps to build the discomfort in order to make it understood by the listener. The analysis carried out on two singular situations shows that listener mobilizes hetero-reformulations which can be self-triggered or hetero-triggered. The appellant, for his part, more specifically appropriates self-triggered self-reformulations. This analysis carried out on poly-callers also reveals that the reformulation is not marked by the clues usually retained to identify it, and that the computer tool can bring out new clues through paradigmatic lists.

Data acquisi-
tion, Knowl-
edge extrac-
tion

Augmented reality HUD vs. conventional HUD to perform a navigation task in a complex driving situation [5] In this study we have investigated the added value of an augmented reality head-up display (AR-HUD) in relation to a conventional head-up display (C-HUD) to perform navigation tasks in a complex road situation. The notion of complexity was defined according to two main factors: infrastructure and traf-

fic. It was used to identify and select real road situations presenting different sources of complexity. This study focuses on one of these situations, which was reproduced on a simulator and broken down into three use cases. A total of 32 participants performed three navigation tasks, using the AR-HUD or the C-HUD. Both objective and subjective data were collected. Data analyses, using linear mixed model analyses of variance and multilevel logistic regression, indicate a slight advantage of the AR-HUD. Participants using the AR-HUD make fewer errors and drive faster on average. Moreover, the AR-HUD is assessed to be more useful and easier to understand than the C-HUD. However, this interface shows limitations, in particular, because it does not enable drivers to anticipate the manoeuvre to be made. The study raises questions about the design of an instrument system that would help drivers not only identify, but also build a representation of a forthcoming manoeuvre to be performed.

Virtual signer - Text to speech French sign language synthesis and translation [6], [12], [11] Our multimodal *SignCom* synthesis system that translates text-to-LSF (French sign language) by means of a 3D virtual character (also called virtual signer) has been completed. This system is based on a multichannel composition mechanism, each channel being associated to linguistic information (from the phonological level to the lexical, syntactic or semantic levels), or to articulatory information. The composition is based on a spatio-temporal arrangement of these elements that are parallelized, and is able to edit and generate utterances in LSF. The new synthesis modules that enrich the initial system are described, including body movement synthesis, facial and hand movement synthesis. They implement grammatical dynamics specific to sign language, based on the fundamental concepts of spatiality and iconicity. This study also details the main technological challenges still to be met in the field of neural machine translation for sign languages.

Generation

Knowledge graph to help create cooking recipes [18] We present a knowledge graph that models cooking recipes to manage human-cobot interaction (collaborative robot) during the preparation of meals. We also present tools for the enrichment and exploration of the graph.

Data acquisition,
Multi-level
representations

Voice Cloning Applied to Voice Disorders [?] Organic dysphonia can lead to vocal impairments. Recording patients' impaired voice could allow them to use voice cloning systems. Voice cloning, being the process of producing speech matching a target speaker voice, given textual input and an audio sample from the speaker, can be used in such a context. However, dysphonic patients may only produce speech with specific or limited phonetic content. Considering a complete voice cloning process, we investigated the relation between the phonetic content, the length of samples and their impact on the output quality and speaker similarity through the use of phonetically limited artificial voices. The analysis of the speakers embedding which are used to capture voices shows an impact of the phonetic content. However, we were not able to observe those variations in the final generated speech.

Generation

Multi-level
representations,
Use cases and
evaluation

Comparative study of lexical embeddings for the extraction of named entities in French [10] We conducted a comparative study of lexical embedding methods for french language on the task of Named Entity Recognition (REN). The objective is to compare the performance of each method on the same task and under the same working conditions. We use the french proportion of the WikiNER corpus as our study corpus. This is a corpus of 3.5 million tokens with 4 types of entities. 10 types of lexical embeddings are studied, including non-contextual embeddings, contextual embeddings and transformer-based embeddings. For each embedding, we train a BiLSTM-CRF as a multi-label classifier to solve the extraction task. For transformer-based models, we also compare their performance when a fine-tuning procedure is used.

Generation

Voice Cloning: Training Speaker Selection in Limited Multi-Speaker Corpus [13] Text-To-Speech synthesis with few data is a challenging task, in particular when choosing the target speaker is not an option. Voice cloning is a popular method to alleviate these issues using only a few minutes of target speech. To do this, the model must first be trained on a large corpus of thousands of hours and hundreds of speakers. We tackled the challenge of cloning voices with a much smaller corpus, using both the speaker adaptation and speaker encoding methods. We studied the impact of selecting our training speakers based on their similarity to the targets. We trained models using only the training speakers closest/farthest to our targets in terms of speaker similarity from a pool of 14 speakers. We showed that the selection of speakers in the training set has an impact on the similarity to the target speaker. The effect is more prominent for speaker encoding than adaptation. However, it remains nuanced when it comes to naturalness.

Knowledge
extraction

Analogical Proportions and Binary Trees [4] Analogical reasoning has been thought for a longtime as something aside, away from logical reasoning. However, it is not exactly so. This chapter in its first part mainly surveys works of the last decade, which have proposed a logical modeling of analogical proportions (i.e., statements of the form "a is to b as c is to d") among other logically expressible proportions and have shown their use in analogical inference. It also emphasizes the pervasiveness of analogical proportions as soon as we compare situations. The second part of the chapter shows that dichotomous trees built from pairs of mutually exclusive properties have also a reading in terms of Boolean analogical proportions, thus providing another clue of the links existing between analogy and logically expressed taxonomies. This also gives birth to noticeable opposition structures and can be related to formal concept analysis.

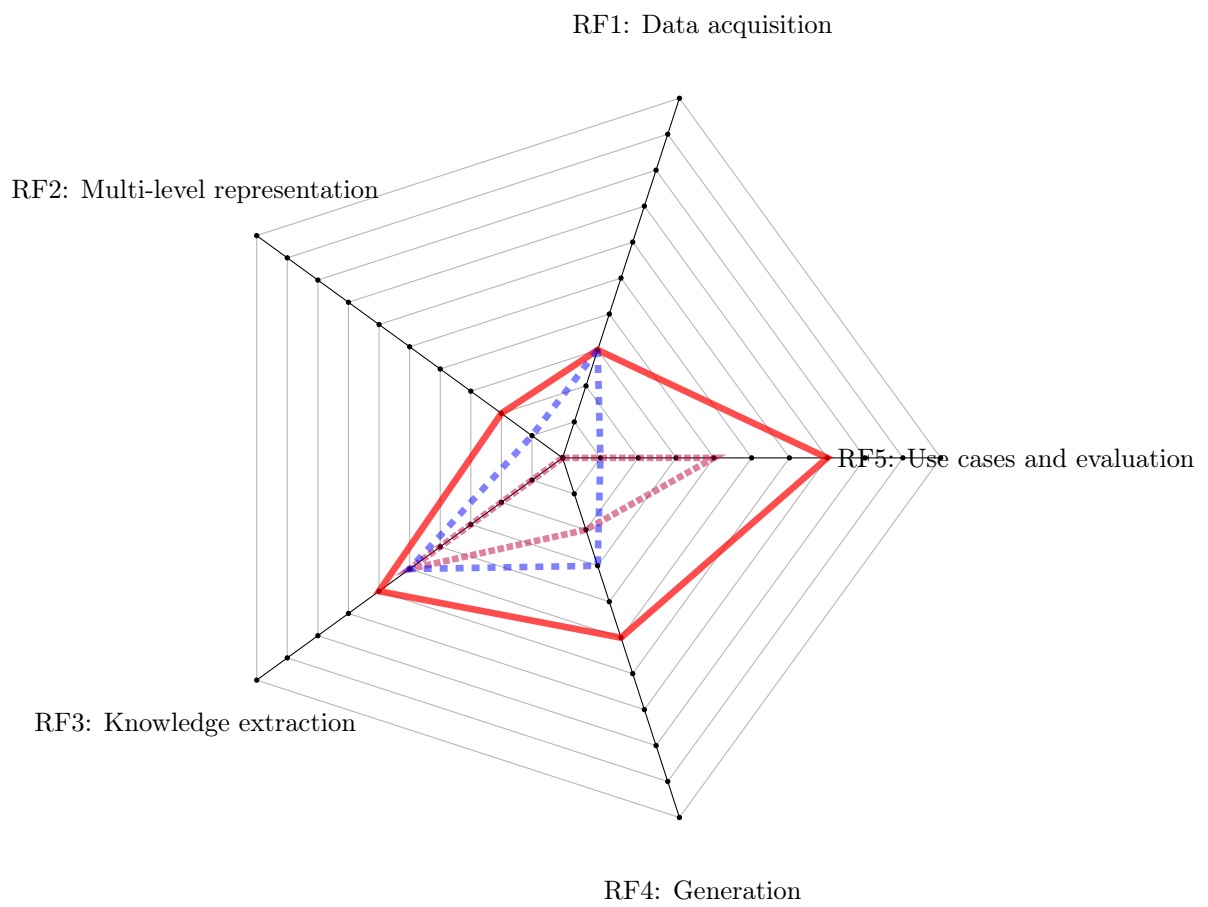


Figure 2: Contributions to each research focus of the team in 2021 (red, solid) compared to 2020 (purple, dotted) and 2019 (blue, dashed)

Summary of the contributions

3.2 Defended PhDs and HDRs

- Alexis Blandin has defended his thesis work [1] is entitled "Opinion analysis for customer relationship optimization" the 27th of March 2023; it was carried out within the framework of a CIFRE agreement, in partnership with the company UNEEK, which offers customer relationship management services. The subject is vast and cuts across many fields, such as databases interrogation in natural language, deep learning, automatic language processing, analysis and detection of emotions or marketing and social sciences. Our work essentially addresses two issues: the querying of tabular data in French and the study of newsletters. In a first part, we explore the possibility of designing a query interface allowing to query in natural language the tabular databases of the company. Then, we explain how and why we have refocused our work on the study of a particular communication channel from the company to the contacts: newsletters. We first study how the emotions transmitted by the text of newsletters influence their perception, and how the results of this study can help in their writing and editing. Then, we propose a modeling of newsletters in the form of heterogeneous graphs, allowing to take into account the visual aspects of text areas and their layout in the newsletter, in addition to their content. We use deep learning techniques such as graph convolution networks, and attention techniques to predict newsletter performance. This original modeling of newsletters produced encouraging results for the prediction task considered. The approach could be further developed in future work to take into account other significant components of newsletters, in particular images. Furthermore, this modelisation could be applied in other multi-modal studies.
- Olivier Zhang defended his PhD on the 21st of December 2023. The subject of this thesis is "Multi-aspect neural analysis and synthesis methods and their application to voice". This thesis was carried out in the framework of a CIFRE contract with Orange. The work presented in this thesis aims to bridge the gap between speech processing and untangling, by exploiting state-of-the-art disentangling models to identify speech attributes automatically, and ultimately improve the control of speech synthesis.

3.3 On going PhDs

1. Mansour Tchégnon is at the end of his PhD. During his third year, he proposed a motion correction system using deep learning techniques and Laplacian motion modeling. The approach consists of using state-of-the-art approaches to first estimate 3D poses sequence and then apply correction to these poses to obtain a motion with better temporal and skeleton coherence. The results are submitted in the international journal Computer & Graphics. He has also generated a dataset of motion captured on people with motor disabilities. A description of this novel dataset has been submitted to the conference Handicap 2024. He is currently writing his PhD manuscript for a defense in the first trimester of 2024.

2. Danrun Cao has started her PhD in the context of an industrial CIFRE contract with Octopus Mind, located in Nantes. Her work addresses information extraction (entities and relationships) in the context of business intelligence, while exploiting multilingual call for tenders. Danrun has published during her first year a conference paper in the ACM DocEng conference, and a conference paper in TALN during her second year. She is currently developing state of the art Deep Learning approaches to Name Entity Recognition and semantic relation extraction for the French language.
3. Clémence Mertz has completed her third PhD year in 2022. She is in her gap year in 2023.
4. Lily Wadoux has completed her third PhD year during 2023. She works on “Voice preservation: adaptation of voice cloning to pathological cases”. During this year, she worked on improving the pipeline for voice cloning in French. The principal issue lies with the availability of large high-quality multi-speaker corpora in French. The conducted experiments focused on automatic selection of training and reference data according to both speaker and quality, as well as new experiments with artificially disordered speech. Two papers were published in conferences [13, 16] and one in a francophone journal [8]. She is currently redacting her PhD.
5. Hoan My Tran started his PhD in November 2022 on the detection of deep speech fakes in the context of voice conversion and speech synthesis. The aim of this work is to propose an automatic voice fraud detection system, based on the state of the art in the field, in particular in comparison with the work proposed as part of the ASVSpooF challenge. A future objective will be to see whether lip movement reading can provide information to improve detection. This PhD is co-directed by Damien Lolive and Pierre-François Marteau, and co-supervised by Arnaud Delhay and Aghilas Sini from LIUM in Le Mans.
6. Ulysse Oliveri started his PhD in March 2023 under a CIFRE contract funded by Airbus. He works on controlled generation of text to fight disinformation. This thesis is directed by Damien Lolive and Bruno Grilhères (Airbus) and co-supervised by Arnaud Delhay, Guillaume Gadek (Airbus) and Benjamin Costé (Airbus).
7. Pauline Mas started her PhD in September 2023 under a CIFRE contract funded by Voxygen. She works on multilingual modulation of prosodic elements in Speech Synthesis. This PhD is directed by Damien Lolive and Olivier Rosec (Voxygen), and co-supervised by Jonathan Chevelu and Marion Ouédraogo (Voxygen).
8. Quentin Lemesle started his PhD in October 2023 on the constraint paraphrase generation task in a frugal approach. This PhD is directed by Damien Lolive, co-supervised by Jonathan Chevelu and Arnaud Delhay. This PhD is half funded by the DGA AID.

4 Software development

MiS The *MiS* corpus has been recorded in June 2022 and has been designed to study the *style* in motion. It consists of a set of walks, a sensory path, jumps and waving movements, in a controlled setting with a choice of participants of various ages (from 8 to 69 years old), sex, weight and size. The originality of this dataset is the style variations on imposed movements and the diversity of subjects. The aim is to use this data to analyse the style in movements and study the foot-floor contacts.

Text-To-Speech system In the frame of several technology transfer projects, developments have been done in 2019 on many pieces of software involved in the team's text-to-speech system to make it usable in industrial environments. This has mainly consisted in shifting some tools from script languages (Python, Perl, shell) to C++, and setting up web services on a production server, compatibility for Android and ARM7 architectures. This has been mainly achieved in the frame of the SPAM project (maturation project funded by the transfer technology service of University of Rennes 1). Developments are going on from then in the frame of different projects such as NADINE or KALYGO-DYS.

Visual Quantum Simulator [Gildas M n ier] Most of the human behaviors involve uncertainty and fuzzy inner decisions in a way or another : Expression can be seen as a individual by-product as the top level visible artifacts of some unknown inner process. Markov models, or statistical process can be used to mimics the observable results of observable behaviors. Quantum State evolution shares also many similar features with the obfuscated or hidden way of thinking and communicating.

Not only does this (relatively) new paradigm bring some ideas related to human decision's process, it also may be a key progress for Machine Learning - Quantum Machine Learning or QML -. Some intensive computation involved in ML may be sped-up using Quantum schemes.

We are investing some efforts in the study of Quantum Computing both as a computing potential asset and source of ideas for human inner process simulation.

A Scala DSL (domain specific) language has been developed to manage Quantum Algorithm Evaluations (<https://github.com/gmenier/VisualQuantumSimulator/wiki>)

The software has been registered by *APP* (Agence pour la Protection des Programmes) under the number Inter Deposit Digital Number IDDN.FR.001.300006.000.S.C.2021.000.10000 (06.56.11046).

This simulator is designed to allow a graphical view of the inner processing involved in Quantum computing, thus helping grasping its inner working.

It is also designed as a starting backbone for experimentation and is actually used as a support for education.

5 Contracts and collaborations

5.1 National Initiatives

5.1.1 TextToKids ANR Project

Participants: Nicolas Béchet, Jonathan Chevelu, Damien Lolive, Alexis Blandin, Rashedur Rahman.

The TextToKids ANR project is the continuation of the homonymous CNRS PEPS project, previously running in 2018. It aims to facilitate the writing and the filtering of texts for children, especially but not only in order to tell them about current events (e.g., presidential elections, Brexit, reception of migrants in France, etc.) in respect of their language skills. The targeted age group is that of young readers, that is, the 7-12 age group. The consortium, which brings together linguists, computer scientists and specialized journalists, will seek to characterize the linguistic constraints to be respected for such a purpose and to propose assistive tools (automated textual analysis, search engine, reformulation, good practices). In terms of benefits, the project works in the direction of a "children's Internet" and opens a way to other modalities (speech, images).

This project is coordinated by Delphine Battistelli (MoDyCo lab), assisted by Nicolas Béchet.

5.1.2 SignToKids ANR Project

Participants: Sylvie Gibet, Clément Reverdy, Damien Lolive, Nicolas Béchet, Vincent Segonnes, Pierre-Francois Marteau, Jeanne Villaneau, Farida Said.

The SignToKids project aims to improve the inclusion of deaf children in the world of education with a focus on digital accessibility. More specifically, it focuses on the creation of the first digital and pedagogical tools to facilitate the coordinated learning of LSF and written French for deaf children. By working on their signed language skills, deaf children will have a more robust native language, making the transition to written French easier. The resulting digital assistance tools (web applications, serious games, or video-books) will serve as a basis to test the skills acquired at different levels of the child's learning. This project will also contribute to the advancement of research on automatic translation from text to signed movements, using recent AI techniques, thus increasing the amount of data available in French and LSF, and ultimately it will enable additional resources and knowledge to be integrated into tools that will be used by deaf children. Practical experiments will be conducted to evaluate these innovative tools with deaf children in LSF-French bilingual schools.

This project is coordinated by Sylvie Gibet. It started in January 2023 and will end at the end of 2025.

Keywords: French Sign Language, Text-to-LSF, pedagogical tools

5.1.3 Breton Synthesis

Participants: Damien Lolive, Antoine Perquin, David Guennec, Gaëlle Vidal.

We conducted experiments aiming to deepen and complement works on Breton language, as a continuation of the project funded by *OPLB - Office Public de la Langue Bretonne*, which led in 2021 and 2022 to build speech resources and a speech synthesis engine based on neural network modeling. In 2023, we finished the post-processing of speech resources recorded at the end of 2022 and finalized the engine and TTS models. Those were subsequently handed over to OPLB. The engine is now accessible on their website at <https://niverel.brezhoneg.bzh/fr/sintezenn/>.

5.1.4 Kaligo+

Participants: Damien Lolive, Simon Giddings, Antoine Perquin, Philippe Martin.

This project aims at proposing a set of digital application on tablets and using virtual reality tools (like Oculus Quest) to favor training children with oral and written troubles. In particular, our role is to work on the oral part of the problem by proposing Text-To-Speech technologies to make an oral feedback to users, and pronunciation analysis tools to develop pronunciation training methods. This project begun in 2022 and is expected to finish in August 2024.

5.1.5 Détection de Voix de Synthèse - DVS

Participants: Damien Lolive, Vincent Barraud, Arnaud Delhay, David Guennec.

The DVS project is a part of the RAPID *Régime d'Appui à l'Innovation Duale* framework proposed by the AID *Agence Innovation Défense*. This framework provides support to innovative projects related to national security and lead with Whispeak. The aim of the DVS project is to design a strategy to detect audio deep fakes. Expression team provides a set of state of the art systems (synthesis, voice conversion and voice cloning), trained or fined tuned with 12 languages with a hundred of speakers each. This project begun in October 2023 and is expected to end in October 2025. Two post-doctoral searchers should be hired by January 2024 for a 18 month duration.

5.1.6 Explicite Voice Attributes ANR Project

Participants: Damien Lolive, Vincent Barraud.

The objective of this project is to crack the codes of human voices via learning explicit and structured representations of voice attributes. The realization of this main objective will have a strong scientific and technological impact, in at least two fields of application: firstly, in speech analysis, it will unlock the understanding of the complex entanglement of the characteristics of a human voice; secondly, in voice generation, it will open the way to a wide range of applications to create a voice with the desired

attributes, allowing the design of so-called voice persona.

The project will leverage a substantial skill set by the partners, which include Orange, a leading telecommunication provider with strong research skills on speech processing, IRCAM, a research lab with a strong expertise in neural speech synthesis and voice conversion and three academic labs: LIA with a strong expertise in speaker analysis and anonymization, LPP, an internationally recognized laboratory, which holds a LABEX (Empirical Foundations of Linguistics), and provides the strong skills needed in phonetics, and IRISA (Expression Team) with expertise in expressive text-to-speech.

The project started in October 2023 and will end in September 2027. The project is registered under the ANR Contract Number ANR-23-CE23-0018.

5.1.7 Koralie - Plan de relance

Participants: Damien Lolive, Antoine Perquin, Aghilas Sini.

This project aims at developing pronunciation analysis and speech synthesis tool for different language in the context of education. This project is funded by *France Relance* and is registered under the record number 2021_000583. it will last 24 months and started the 1st December 2021.

5.1.8 Ethos - Automatic annotation of Horse Sound

Participants: Philippe Martin, Arnaud Delhay, Vincent Barraud.

In June 2023, the team CIRCÉ² from EthoS laboratory³ and the team EXPRESSION began collaborating on the detection and classification of sound events emitted by horses. Horses are equipped with microphones to pick up various sound events linked to their behavior (neighing, snorting, etc.) in different environments (stable stalls, outdoors). In this context, classification work enables ethology researchers to identify events in the sound continuum that may occur, and then to observe the scenes linked to these events, thus studying the well-being of the subject observed as a function of what happens in these situations.

The team CIRCÉ has recorded about 30 000 hours of sound rush from 65 horses in diverse location as meadow or stable. Each horse was recorded for 2 continuous weeks. There is a wide range of sounds with their particularities and also rarer events which make this task a complex one.

The main objective of this collaboration will be to research methods for adapting to different types of massive acoustic content, based on recordings made continuously, 24 hours a day, 7 days a week, on animals equipped with on-board loggers fitted with sensors (audio recorder and accelerometer). The aim is to develop versatile classification methods using unsupervised or weakly supervised classification methodologies, and to evaluate the ability to transfer knowledge from one model to another on ethological data.

²Communication Intra-/Inter espèces, Relation, Cognition, Émotions

³UMR 6552 - Ethologie animale et humaine

5.2 National Collaborations

- We are collaborating actively with Marie Tahon, associate professor at LIUM in Le Mans, France. We shared common interests on emotions annotation, and generation in the context of speech synthesis.
- In the frame of the TREMoLo and TextToKids projects, the team works with members of the MoDyCo lab, especially Delphine Battistelli, full professor at University of Paris-Nanterre. 2 MSc interns have been co-supervised (Alexis Blandin, Aline Étienne), while Jade Mekki defended her PhD in September 2022.
- Nelly Barbot and Aghilas Sini have collaborated with Laurent Miclet (former) emeritus professor at University of Rennes 1 about the supervised classification based on analogical proportions.
- Caroline Larboulette collaborates with Laura Pouppeville, an artist from the "ateliers Montebello" in Lyon, on the project Motion in Style, that has been awarded a grant from the IRISA project "1 artiste - 1 chercheur". The aim of the project, that started in december 2021, is to collect and analyse data, and use this data to produce an art piece that shall be presented at the "fête de la science" or similar venues. They collected data in june 2022. The data is still being processed and analyzed.
- Since december 2021, Caroline Larboulette collaborates with Paul Richard (associate professor at Polytech' Angers, LARIS lab) and Jérémy Besnard (associate professor at Université d'Angers, laboratoire de Psychologie des Pays de la Loire) on the capture and synthesis of avatars, for the study of social interactions in Virtual Reality.

5.3 International Collaborations

- In 2017, we have developed a collaboration with Ingmar Steiner and Sébastien Le Maguer from Saarland University, Saarbruck, Germany. Notably, we recruited an internship to work together on the construction of a common interface for Speech synthesis systems enabling to visualize and interact with several systems, like Expression TTS systems and also MaryTTS. We have continued this collaboration since 2018 with Sébastien Le Maguer, who has moved to the ADAPT center, Trinity College Dublin.
- The collaboration started in 2018 with John D. Kelleher (TU Dublin) has continued with the PhD of Somaye Jafaritazehjani, co-supervised by Damien Lolive and Gwénoél Lecorvé.
- Since September 2022, Sylvie Gibet collaborates with Sofia Dahl (Aalborg University, Denmark) and Doga Cavdir (postdoctorate) on expressive musical gestures.
- Since October 2021, Caroline Larboulette collaborates with Ravi Dattatreya, researcher and Managing Partner of Neonyx Technology, New York, and a master student from ENS Rennes, Baptiste Demoussel, on the String Art project. Since

October 2022, they collaborate on a new Art and Science project, still dealing with Computational Art: the Kolam project. The aim is to create a unified/general algorithm capable of drawing Indian Kolams and Celtic Knots using the same formalism. They work with a new student, Hugo Boulier, from ENS Rennes as well.

6 Dissemination

6.1 Involvement in the Scientific Community

- Sylvie Gibet is a member of various program committees for international conferences: IEEE International Conference on Acoustics, Speech, and Signal Processing, Motion in Game, Computer Animation and Social Agents, Motion Computing. She serves as a reviewer for various journals (Computer & Graphics, Trans. on Visualization and Computer Graphics, Computer Animation and Virtual Worlds, The Visual Computer, Springer Nature, Universal Access in the Information Society). In 2023, she was a reviewer for a Carnot project. She participated to the PhD committee of Léon Victor in Univ. Lyon1, and was the president of the PhD committees of Alberto Jovane at Univ. Rennes1 (Feb. 2023) and of Ariel Kwiatkowski at École Polytechnique in Paris-Saclay (Nov. 2023).
- Giuseppe Berio served as reviewer for national and international conferences. He also acts as steering committee member of the International Conference on Advances in Semantic Processing (SEMAPRO).
- Jonathan Chevelu is an elected member of the 'Conseil scientifique' (Research committee) of École Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT).
- Nelly Barbot has served as a reviewer for the International conference of the International Speech Communication Association (Interspeech 2023) and the IJCAI-ECAI 2023 workshop on Interactions between Analogical Reasoning and Machine Learning. She is an elected member of the 'Conseil scientifique' (Research committee) of ENSSAT.
- Arnaud Delhay is an elected member of the 'Commission Recherche' (Research committee) of the IUT of Lannion. He has served as a reviewer for Joint International Conference on Language Resources and Evaluation (LREC-COLING 2024), and the International conference of the International Speech Communication Association (Interspeech 2023).
- Caroline Larboulette is a member of various program committees for international conferences, a member of the editorial review board of the international journal of computer graphics and creative interfaces (IJCICG) and serves as a reviewer for various journals (Computer & Graphics, TVCG, CAVW, Journal on Multimodal User Interfaces). In 2021, she joined the Editorial Board of Frontiers in Computer Science as an Associate Editor, to participate to the launch of a new section on Computer Graphics and Visualization in 2022. Since november 2019, she is

co-director of the GTAS, the "Groupe de Travail Animation et Simulation" of the GdR IG-RV (Informatique Géométrique et Graphique, Réalité Virtuelle et Visualisation) of the CNRS INS2i and also member of the direction committee of the GdR.

- Damien Lolive is head of the Signal, Image, Language Department (formerly Media and Interaction Department) of IRISA lab from October 2021. He is part of board of directors of the French speech communication association (AFCP). He is also co-responsible for the working group on inter-modality and multi-modality of the GDR TAL (special interest group in NLP). He serves as a reviewer for the IEEE Transactions on Speech and Language processing, for the IEEE Access journal, for the *Traitement Automatique des Langues* journal, for the International conference of the International Speech Communication Association (Interspeech), the International Conference on Audio, Speech and Signal Processing (ICASSP), the international conferences LREC and Speech Prosody, as well as for the *Journées d'Études sur la Parole* conference. He regularly serves as an expert for the french research agency (ANR) and also for the Ministry of Research to expertise CIR application inf the NLP domain. He also participated to 6 Phd committees during 2023: 4 as a reviewer (Sondes Abderrazek, Clément Le Moine, Brooke Stephenson, Vinicius Ribeiro) and 2 as examiner and president of the jury (Antoine Chaffin, Rami Rigal).
- Nicolas Béchet is a member of program committees for international conferences International Conference on Natural Language & Information Systems (NLDB) and Language Resources and Evaluation Conference (LREC). He also has served as a reviewer for the Intelligent Decision Technologies Journal (IDT), the International Conference on Computational Linguistics (COLING) and the International Journal of Machine Learning and Cybernetics.
- David Guennec served as a reviewer for the International conference of the International Speech Communication Association (Interspeech 2023) and the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2024).

6.2 Teaching

- Giuseppe Berio teaches courses on design and implementation of decision information systems at IUT Vannes and faculty of sciences (master level), Université Bretagne Sud. A project management course is also taught at IUT Vannes.
- Nelly Barbot teaches the following mathematics courses at École Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT): algebra and analysis basis, mathematical logic in Licence level, probability and statistics in Master level. She is responsible of the student admission processes at ENSSAT.
- Vincent Barraud teaches the following computer science courses at École Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT): Web Programming, Web Services and Distributed Algorithms in Licence level. He has

been director of studies at ENSSAT until mid 2021 and is now responsible for the master 1 level of computer science.

- Nicolas Béchet teaches various computer sciences courses at the Statistique et Informatique Décisionnelle department of IUT Vannes. He is director of studies at IUT in STID department.
- Arnaud Delhay teaches databases and web programming (server- and client-side) in Licence levels at IUT of Lannion, calculability and computational complexity of problems in Master level at École Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT).
- Jonathan Chevelu teaches the following computer science courses at École Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT): cybersecurity in Licence and Master level, operative systems in Licence level and natural language processing in Master level.
- Sylvie Gibet teaches the following Computer Science courses at the faculty of sciences, Université Bretagne Sud: An introduction of Digital Signal Processing and Machine Learning (1st year master level, AIDN), and Movement and Artificial Intelligence (2nd year master level, AIDN). She is responsible for the 2nd year of the Master in Computer Science (AIDN program).
- Caroline Larboulette teaches graphical user interfaces to undergraduate students as well as introduction to computer graphics (I2G) and simulation and interactive applications (SAI) at the master level (Master of Computer Science, AIDN (Interactive Applications and Data)) at the faculty of science of the Université Bretagne Sud. She also teaches a character animation lecture at PolyTech' Angers. Since october 2020, she has been deputy head of the MIS (Mathematics - Computer Science - Statistics) department at Université Bretagne Sud. She has been elected Head of this department in november 2022. She is also responsible of both Computer Science degrees, bachelor and master.
- Damien Lolive teaches the following computer science courses at École Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT): object-oriented programming in Licence level, compilers architecture and formal languages theory in Master level, speech and language processing in Master level, and pattern recognition in Master level. Damien Lolive is currently the Head of the Computer Science department at ENSSAT.
- Pierre-François Marteau teaches programming languages, logics, introduction to cryptography, information retrieval, machine learning and intrusion detection courses in computer sciences License and Master levels, mostly at École Nationale Supérieure de Bretagne Sud (ENSIBS). He is responsible of the bachelor level for the computer science program at ENSIBS.
- Gildas Ménier teaches various computer sciences courses at the faculty of sciences, Université de Bretagne Sud.
- David Guennec teaches algorithmics, object oriented programming, machine learning and modelization courses (license and master levels) at ENSSAT.

- Lily Wadoux taught algorithmics, object-oriented programming and operative systems at ENSSAT, then algorithmics, object-oriented programming and databases (licence level) at IUT of Lannion.

6.3 Conferences, workshops, invitations

- Damien Lolive has been member of the organizing committee of the 12th Speech Synthesis Workshop (SSW23). At this occasion, he has been in charge of sponsoring.
- Lily Wadoux helped on-site for the organisation of the 12th Speech Synthesis Workshop (SSW23).

6.4 Meetings, other dissemination

- Lily Wadoux, Aghilas Sini, Antoine Perquin, David Guennec presented a popular science talk about sound recording, speech synthesis & voice cloning (December 10, 2022).
- David Guennec presented an introduction to voice cloning at ENSSAT as part of Conf'Express, a cycle of small format scientific conferences organized by JES (March 14, 2023).
- Sylvie Gibet organized and led the motion capture workshop at Université Bretagne Sud's at the "Nuits de la Science" on October 27, 2023.
- Sylvie Gibet presented the lecture "Gesture as a Means of Expression and Language" at the "Nuits de la Science" 2023 at UBS.

6.5 Graduate student and student internship

- Yasser El Ayyachy, a Master 1 student in computer science at ENSSAT, worked on building a classifier for the detection of vocoded speech. The internship lasted 8-weeks. He was supervised by Aghilas Sini.
- Mathias Peuch, a Master 1 student in computer science at ENSSAT, worked on the improvement of the team' evaluation software Flexeval during a part-time 8-weeks internship. An automatized process for analyzing the results of Flexeval subjective tests was created. He was supervised by David Guennec.
- Clément Delrieux, a Master 1 student in computer science at ENSSAT, has done a 8-weeks internship. He worked on a video annotation platform to build a corpus on anomaly detection, called EMO&LY. He also worked on a speaker detection by watching lip movements. He was supervised by Aghilas Sini and Arnaud Delhay.

7 Bibliography

Doctoral dissertations and “Habilitation” theses

- [1] A. BLANDIN, *Analyse d’opinion pour l’optimisation de la relation client : optimisation de services de gestion de communautés et d’outils de communication à destination du client.*, Theses, Université Bretagne Sud, March 2023, <https://hal.science/tel-04222326>.
- [2] S. JAFARITAZEHIJANI, *Towards an improved understanding of the concept of style and its implications for textual style transfer*, Theses, Technological University Dublin, December 2023, <https://arrow.tudublin.ie/sciendoc/270/>.
- [3] O. ZHANG, *Méthodes neuronales pour le traitement de la parole : vers le démantèlement des attributs de la voix*, Theses, Université de Rennes, December 2023, <https://www.theses.fr/s248315>.

Articles in referred journals and book chapters

- [4] N. BARBOT, L. MICLET, H. PRADE, G. RICHARD, “Analogical Proportions and Binary Trees”, in: *Logic in Question: Talks from the Annual Sorbonne Logic Workshop (2011-2019)*, B. Jean-Yves, D. Jean-Pierre, M. Amirouche, and P. A. Christine (editors), *Studies in Universal Logic book series (SUL)*, Springer International Publishing, January 2023, p. 435–458, <https://inria.hal.science/hal-04147804>.
- [5] C. CHAUVIN, F. SAID, S. LANGLOIS, “Augmented reality HUD vs. conventional HUD to perform a navigation task in a complex driving situation”, *Cognition, Technology and Work Volume 25*, Issue 2-3, 2023, <https://hal.science/hal-04055887>.
- [6] S. GIBET, “Avatar signeur – Synthèse de la langue des signes française à partir de texte [Virtual signer - Text to speech French sign language synthesis]”, *Revue TAL : Traitement Automatique des Langues 63*, 2, 2023, p. 67–91, <https://inria.hal.science/hal-04448954>.
- [7] F. PUGNIERE-SAAVEDRA, N. BÉCHET, N. GARRIC, G. LEDEGEN, R. KESSLER, “Place, rôle(s) et forme(s) de la reformulation dans la relation d’aide d’un webchat”, *Travaux de linguistique du CERLICO*, 34, 2023, p. 121–141, <https://hal.science/hal-04278228>.
- [8] A. SINI, L. WADOUX, A. PERQUIN, G. VIDAL, D. GUENNEC, D. LOLIVE, P. ALAIN, N. BARBOT, J. CHEVELU, A. DELHAY, “Techniques de synthèse vocale neuronale à l’épreuve des données d’apprentissage non dédiées : les livres audio amateurs en français”, *Revue TAL : traitement automatique des langues 63*, 2, June 2023, p. 41–65, <https://inria.hal.science/hal-04139701>.

Publications in Conferences and Workshops

- [9] N. BÉCHET, R. KESSLER, G. ICART, G. LEDEGEN, “Construction d’un corpus annoté en genre par apprentissage zero-shot”, in: *Extraction et Gestion des Connaissances*, Lyon, France, January 2023, <https://hal.science/hal-04452065>.
- [10] D. CAO, N. BÉCHET, P.-F. MARTEAU, “Étude comparative des plongements lexicaux pour l’extraction d’entités nommées en français”, in: *18e Conférence en Recherche*

- d'Information et Applications – 16e Rencontres Jeunes Chercheurs en RI – 30e Conférence sur le Traitement Automatique des Langues Naturelles – 25e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues*, C. Servan, A. Vilnat (editors), ATALA, p. 94–104, Paris, France, June 2023, <https://hal.science/hal-04130131>.
- [11] S. GIBET, P. MARTEAU, “Signing Avatars - Multimodal Challenges for Text-to-sign Generation”, in: *17th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2023, Waikoloa Beach, HI, USA, January*, IEEE, p. 1–8, 2023.
- [12] S. GIBET, P. MARTEAU, “A Text-To-SL Synthesis System Using 3D Avatar Technology”, in: *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2023 - Workshops, Rhodes Island, Greece, June 4-10, 2023*, IEEE, p. 1–5, 2023.
- [13] D. GUENNEC, L. WADOUX, A. SINI, N. BARBOT, D. LOLIVE, “Voice Cloning: Training Speaker Selection with Limited Multi-Speaker Corpus”, in: *12th ISCA Speech Synthesis Workshop (SSW2023)*, ISCA, p. 170–176, Grenoble, France, August 2023, <https://inria.hal.science/hal-04223354>.
- [14] S. JAFARITAZEHJANI, G. LECORVÉ, D. LOLIVE, J. D. KELLEHER, “Local or Global: The Variation in the Encoding of Style Across Sentiment and Formality”, in: *International Conference on Artificial Neural Networks*, Springer, p. 492–504, 2023.
- [15] A. SINI, A. PERQUIN, D. LOLIVE, A. DELHAY, “Phone-Level Pronunciation Scoring for L1 Using Weighted-Dynamic Time Warping”, in: *IEEE SLT 2022 - IEEE Spoken Language Technology Workshop*, Doha, Qatar, January 2023, <https://hal.science/hal-03976592>.
- [16] L. WADOUX, N. BARBOT, J. CHEVELU, D. LOLIVE, “Voice Cloning for Voice Disorders: Impact of Phonetic Content”, in: *TSD 2023: Text, Speech, and Dialogue, Lecture Notes in Computer Science, 14102*, Springer Nature Switzerland, p. 293–303, Pilsen, Czech Republic, September 2023, <https://inria.hal.science/hal-04223342>.
- [17] O. ZHANG, O. LE BLOUCH, N. GENGEMBRE, D. LOLIVE, “An extension of disentanglement metrics and its application to voice”, in: *INTERSPEECH 2023*, ISCA, p. 2878–2882, Dublin, Ireland, August 2023, <https://hal.science/hal-04449163>.

Miscellaneous

- [18] F. SAÏD, J. VILLANEAU, S. BENFERHAT, A. BIGER, T. CELESTE, V. CADOU, K. PHILIPPE, “Graphe de connaissances pour l’aide à la réalisation de recettes de cuisine”, EGC’ 2023, *RNTI-E-39*, January 2023, Poster, <https://hal.science/hal-03976515>.