UMR IRISA

Activity Report 2018

Team EXPRESSION

Expressiveness in Human Centered Data/Media

D6 – Media and Interactions

1   Team composition

1.1   Composition

Head of the team

    Pierre-François Marteau, Professor, Université Bretagne Sud

Administrative assistants

    Sylviane Boisadan, Université Bretagne Sud
    Angélique Le Pennec, Université de Rennes 1
    Joëlle Thépault, Université de Rennes 1

Permanent members

    Nelly Barbot, 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-Lorrain, Associate professor, Université de Rennes 1
    Sylvie Gibet, Professor, Université Bretagne Sud
    Caroline Larboulette, Associate professor, Université Bretagne Sud
    Gwénolé Lecorvé, Associate professor, Université de Rennes 1
    Damien Lolive, Associate professor, Université de Rennes 1
    Gildas Ménier, Associate professor, Université Bretagne Sud
    Jeanne Villaneau, Associate professor (emeritus), Université Bretagne Sud

Associate members

    Vincent Barreaud, Associate professor, Université de Rennes 1
    Elisabeth Delais-Roussarie, Senior researcher, CNRS/LLF
    Jean-François Kamp, Associate professor, Université Bretagne Sud
    Farida Said, Associate professor, Université Bretagne Sud

Non-permanent members

    Rémy Kessler, Post-doctoral researcher, Université de Bretagne Sud (since December 2016)
→ Waseem Safi, Post-doctoral researcher, Université de Rennes 1 (from November 2018)
→ Quentin Di-Fant, Engineer, Université de Rennes 1 (from November 2018)
→ Somayeh Jafaritazehjani, Engineer, Université de Rennes 1 (from November 2018)
→ Cédric Fayet, Université de Rennes 1, ATER (from September 2018)

PhD students

      Lei Chen, Université de Bretagne Sud/Univ. McGill, ARED, 6th year

      Cédric Fayet, Université de Rennes 1, DGA/ARED, defended on December 2018

      Lucie Naert, Université Bretagne Sud, CDE, 3rd year

      Nicolas Bloyet, Université de Bretagne Sud, Thèse CIFRE Seed, 2nd year

      Stefania Pecóre, Université de Bretagne Sud, 3rd year

      Antoine Perquin, Université de Rennes 1, CD INSA, 1st year

      Clément Reverdy, Université Bretagne Sud, CDE+ANR InGredible, final year

      Meysam Shamsi, Université de Rennes 1, ARED/CD22, 2nd year

      Aghilas Sini, Université de Rennes 1, LABEX EFL/ANR SynPaFlex, 2nd year

      Ahmad Alaa Eddine, Université Bretagne Sud/Lebanese University, CIOES, 3rd year

 → Valentin Durand De Gevigney, Université Bretagne Sud, DGA/CDE, 1st year

 → Betty Fabre, Université de Rennes 1, Thèse CIFRE Orange Labs, 1st year

 → Tiago Brizolara, Université Bretagne Sud, ARED/CD56, 1st year

 → Jade Mekki, Université de Rennes 1, ANR TREMoLo, 1st year

Master students

 ⇄ Henri Lasselin, Université de Rennes 1, ENSSAT

 ⇄ Frédéric Le Bellour, Université de Rennes 1, ENSSAT

1.2 Evolution of the staff

The permanent staff has been stable during the year. The number of PhD students is increasing with 1 PhD defense and four newly hired PhD students: Betty Fabre, Valentin Durand de Gevigney, Jade Mekki and Tiago Brizolara. 3 non-permanent members arrived on November 2018: Waseem Safi and Quentin Di-Fant have been hired in the frame of 2 new projects on text-to-speech synthesis, and Somayeh Jafaritazehjani started as an engineer prior to a recruitement as a PhD candidate on 1st January 2019.

The PhD defense of Cédric Fayet was held on 18th December 2018. Cédric Fayet holds since September 2018 a teaching-research position (ATER) at ENSSAT Lannion.

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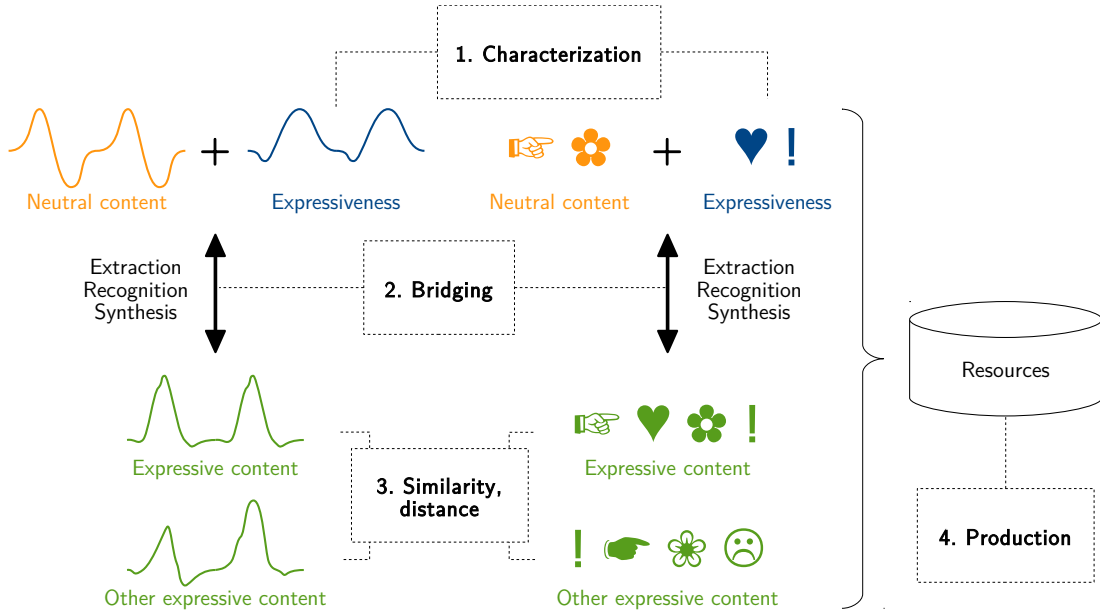
Figure 1: Overview of the main challenges considering both on continuous numerical (left) and discrete symbolic (right) data.

(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:

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 embeddings[1] (metric or similarity space embeddings) in order to ease post-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.

---

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

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 <u>second challenge</u>;

3 Scientific achievements

3.1 New Results by Key Issues

In accordance with the Team Project, the main outcomes for 2018 are listed into the following key issues items defined above for the team:

Data acquisition

LSF-Animal Corpus: A new LSF corpus and associated motion captured (mocap) dataset has been captured in may 2018. The mocap data includes full body, hands and simplified facial data in French Sign Language performed by a deaf male adult, professor of LSF.

Data acquisition

LSF Motion Capture Corpora for Signing Avatars [9]: This paper describes four corpora that have been designed and built in our research team. These corpora have been recorded using motion capture (MoCap) and video equipment, and annotated according to multi-tiers linguistic templates. Each corpus has been designed for a specific linguistic purpose and is dedicated to data-driven synthesis, by (i) replacing signs or groups of signs within an utterance, (ii) replacing phonetic or phonological components and in this way modifying the grammatical or semantic aspects of the phrase, or (iii) altering prosody in the produced sign language utterances.

Data acquisition

CONDUCT: An Expressive Conducting Gesture Dataset for Sound Control [7]: Recent research in music-gesture relationship has paid more attention on the sound variations and its corresponding gesture expressiveness. In this paper we are interested by gestures performed by orchestral conductors, with a focus on the expressive gestures made by the non dominant hand. We make the assumption that these gestures convey some meaning shared by most of conductors, and that they implicitly correspond to sound effects which can be encoded in musical scores. Following this hypothesis, we defined a collection of gestures for musical direction. These gestures are designed to correspond to well known functional effect on sounds, and they can be modulated to vary this effect by simply modifying one of their structural component (hand movement or hand shape). This paper presents the design of the gesture and sound sets and the protocol that has led to the database construction. The relevant musical excerpts and the related expressive gestures have been first defined by one expert musician. The gestures were then recorded through motion capture by two non experts who performed them along with recorded music. This database will serve as a basis for training gesture recognition system for live sound control and modulation.

***SynPaFlex* speech corpus [20]:**   This paper presents an expressive French audio-books corpus containing eighty seven hours of good audio quality speech, recorded by a single amateur speaker reading audiobooks of different literary genres. This corpus departs from existing corpora collected from audiobooks since they usually provide a few hours of mono-genre and multi-speaker speech. The motivation for setting up such a corpus is to explore expressiveness from different perspectives, such as discourse styles, prosody, and pronunciation, and using different levels of analysis (syllable, prosodic and lexical words, prosodic and syntactic phrases, utterance or paragraph). This will allow developing models to better control expressiveness in speech synthesis, and to adapt pronunciation and prosody to specific discourse settings (changes in discourse perspectives, indirect vs. direct styles, etc.). To this end, the corpus has been annotated automatically and provides information as phone labels, phone boundaries, syllables, words or morpho-syntactic tagging. Moreover, a significant part of the corpus has also been annotated manually to encode direct/indirect speech information and emotional content. The corpus is already usable for studies on prosody and TTS purposes and is available to the community.

Data acquisition

***TED* corpus [3]:**   The European "Tenders Electronic Daily" (TED) is a large source of semi-structured and multilingual data that is very valuable to the Natural Language Processing community. This data sets can effectively be used to address complex machine translation, multilingual terminology extraction, text-mining, or to benchmark information retrieval systems. Despite of the services offered by the user-friendliness of the web site that is made available to the public to access the publishing of the EU call for tenders, collecting and managing such kind of data is a great burden and consumes a lot of time and computing resources. This could explain why such a resource is not very (if any) exploited today by computer scientists or engineers in NLP. We provided two documented and easy-to-use multilingual corpora (one of them is a parallel corpus), extracted from the TED web source that we will release for the benefit of the NLP community.

Data acquisition

***EMO&LY* corpus [8]:**   This paper presents a new corpus, called EMOLY (EMOtion and AnomaLY), composed of speech and facial video records of subjects that contains controlled anomalies. As far as we know, to study the problem of anomaly detection in discourse by using machine learning classification techniques, no such corpus exists or is available to the community. In EMOLY, each subject is recorded three times in a recording studio, by filming his/her face and recording his/her voice with a HiFi microphone. Anomalies in discourse are induced or acted. At this time, about $8,65$ hours of usable audiovisual recording on which we have tested classical classification techniques (GMM or One Class-SVM plus threshold classifier) are available. Results confirm the usability of the anomaly induction mechanism to produce anomalies in discourse and also the usability of the corpus to improve detection techniques.

Data acquisition

Data acquisition, knowledge extrac-

Corpus construction for text classification: This dataset, dedicated to text classification, is composed of RSS feeds items belonging to 6 categories (ART_CULTURE, ECONOMIE/ECONOMY, POLITIQUE/POLITICS, SANTE_MEDECINE/HEALTH_MEDICINE, SCIENCE, SPORT) in French and English languages. The dataset has been collected during November 2018. It is available at `https://github.com/pfmarteau/RSS-Feed-6C-dataset`.

Use cases and evaluation

International Blizzard challenge: We participated for the fourth time to the challenge this year. The process followed to build the voices from given data and the architecture of our system is described in [4]. The system is a concatenative system and uses a selection cost which integrates notably a DNN-based embedding prediction.

Generation

Perceptual Validation for the Generation of Expressive Movements from End-Effector Trajectories [1]: Endowing animated virtual characters with emotionally expressive behaviors is paramount to improve the quality of the interactions between humans and virtual characters. Full-body motion, in particular its subtle kinematic variations, represents an effective way of conveying emotionally expressive content. However, before synthesizing expressive full-body movements, it is necessary to identify and understand what qualities of human motion are salient to the perception of emotions and how these qualities can be exploited to generate novel and equally expressive full-body movements. Based on previous studies, we argue that it is possible to perceive and generate expressive full-body movements from a limited set of joint trajectories. Hence, these selected trajectories define a significant and reduced motion space which is adequate for the characterization of the expressive qualities of human motion and that is both suitable for the analysis and generation of emotionally expressive full-body movements. The purpose and main contribution of this work is the methodological framework we defined and used to assess the validity and applicability of the selected trajectories for the perception and generation of expressive full-body movements. This framework consists of the creation of a motion capture database of expressive theatrical movements, the development of a motion synthesis system based on trajectories re-played or re-sampled and inverse kinematics, and two perceptual studies.

Generation

Automatic disfluency insertion towards spontaneous TTS - formalization and proof of concept [18]: This work is an exploratory work on the automatic insertion of disfluencies in text-to-speech systems. By inserting pauses, repetitions and revisions, the objective is to make synthetic speech more spontaneous and expressive. To achieve this task, we formalize the problem as a theoretical process, where transformation functions are iteratively composed. This is a novel contribution since most of the previous work either focus on the detection or cleaning of disfluencies in speech transcripts, or solely concentrate on pause insertion in text-to-speech. We present a first implementation of the proposed process using conditional random fields and language models, before conducting objective and perceptual evaluations. These experiments lead to the conclusion that our proposition is effective to generate disfluencies, and highlights per-

pectives for future improvements. This work has also been published in TALN'2017 and received a best paper award.

Phone-Level Embeddings for Unit Selection Speech Synthesis [17]: Deep neural networks have become the state of the art in speech synthesis. They have been used to directly predict signal parameters or provide unsupervised speech segment descriptions through embeddings. In this paper, we present four models with two of them enabling us to extract phone-level embeddings for unit selection speech synthesis. Three of the models rely on a feed-forward DNN, the last one on an LSTM. The resulting embeddings enable replacing usual expert-based target costs by an euclidean distance in the embedding space. This work is conducted on a French corpus of an 11 hours audiobook. Perceptual tests show the produced speech is preferred over a unit selection method where the target cost is defined by an expert. They also show that the embeddings are general enough to be used for different speech styles without quality loss. Furthermore, objective measures and a perceptual test on statistical parametric speech synthesis show that our models perform comparably to state-of-the-art models for parametric signal generation, in spite of necessary simplifications, namely late time integration and information compression

Automatic of discourse types in audiobooks [19]: To synthesize audiobooks in an expressive manner, it is necessary to know the type of discourses that have to be produced. However, in a novel or a tale, narrative perspectives and discourse types often change, moving from narrative and recitative paragraphs to direct speech, reported speech, and even dialogs. In this work, we will present a tool that was developed from the analysis of a corpus (including excerpts from Madame Bovary and Les MystÃ¨res de Paris) and that relies on paragraph as basic unit. It allows not only to automatically determine the type of speech (narrative speech, direct speech, dialogs), and therefore to know who is speaking, but also to annotate the extension of the discursive modifications. This later point is important, especially in the case of parentheticals with reporting verbs where the narrator speaks again in the middle of a direct speech sequence. In its current form, the tool achieves a 89% detection rate.

Per Channel Automatic Annotation of Sign Language Motion Capture Data [15]: Manual annotation is an expensive and time consuming task partly due to the high number of linguistic channels that usually compose sign language data. In this paper, we propose to automatize the annotation of sign language motion capture data by processing each channel separately. Motion features (such as distances between joints or facial descriptors) that take advantage of the 3D nature of motion capture data and the specificity of the channel are computed in order to (i) segment and (ii) label the sign language data. Two methods of automatic annotation of French Sign Language utterances using similar processes are developed. The first one describes the automatic annotation of thirty-two hand configurations while the second method describes the

annotation of facial expressions using a closed vocabulary of seven expressions. Results for the two methods are then presented and discussed.

Multi-level represen-tation, Knowledge extraction

Discourse phrases classification: direct vs. narrative audio speech [22]: In the field of storytelling, speech synthesis is trying to move from a neutral machine-like to an expressive voice. For parametric and unit-selection systems, building new features or cost functions is necessary to allow a better expressivity control. The present article investigates the classification task between direct and narrative discourse phrases to build a new expressivity score. Different models are trained on different speech units (syllable, word and discourse phrases) from an audiobook with 3 sets of features. Classification experiments are conducted on the Blizzard corpus which features children English audiobooks and contains various characters and emotional states. The experiments show that the fusion of SVM classifiers trained with different prosodic and phonologic feature sets increases the classification rate from 67.4% with 14 prosodic features to 71.8% with the 3 merged sets. Also the addition of a decision threshold achieves promising results for expressive speech synthesis according to the strength of the constraint required on expressivity: 71.8% with 100% of the words, 79.9% with 50% and 82.6% with 25%.

Generation, Knowledge extraction

Can we Generate Emotional Pronunciations for Expressive Speech Synthesis? [2]: In the field of expressive speech synthesis, a lot of work has been conducted on suprasegmental prosodic features while few has been done on pronunciation variants. However, prosody is highly related to the sequence of phonemes to be expressed. This article raises two issues in the generation of emotional pronunciations for TTS systems. The first issue consists in designing an automatic pronunciation generation method from text, while the second issue addresses the very existence of emotional pronunciations through experiments conducted on emotional speech. To do so, an innovative pronunciation adaptation method which automatically adapts canonical phonemes first to those labeled in the corpus used to create a synthetic voice, then to those labeled in an expressive corpus, is presented. This method consists in training conditional random fields pronunciation models with prosodic, linguistic, phonological and articulatory features. The analysis of emotional pronunciations reveals strong dependencies between prosody and phoneme assimilation or elisions. According to perceptual tests, the double adaptation allows to synthesize expressive speech samples of good quality, but emotion-specific pronunciations are too subtle to be perceived by testers.

Generation, Uses cases and evalua-tion

Disfluency Generation [18, 23, 24]: The synthesis of spontaneous natural speech is a challenge to be met. One way to approach it is to insert disfluencies into a fluent text to be synthesized. This topic already addressed in 2018 has been published [18], continued through the master internship of Henri Lasselin, and extended by Gwénolé Lecorvé and Claude Simon. We have been experimenting with several sequence-to-sequence neural models to transform fluent texts into disfluent ones. Moreover, we expanded a previous work to insert revisions within a fluent text. This second work will be submitted in 2019.

Modelling of language registers [14, 13]:   Language registers are an observable
stylistic trait of texts and speeches. However, they are still poorly studied in natural
language processing. In 2019, we have been studying useful features from linguistics
that could be used for automatic processings, and a method has been proposed to
automatically build a corpus labelled in language registers.

Knowledge
extraction,
data ac-
quisition,
multi-level
representa-
tion

Extraction of terminology in the field of construction: We describe a corpus
analysis method to extract terminology from a collection of technical specifications book
in the field of construction. Using statistics and word n-grams analyzes, we extract the
terminology of the domain and then perform pruning steps with linguistic patterns
and internet queries to improve the quality of the final terminology. In this work, we
specifically focus on the improvements got by applying Internet queries and patterns.
These improvements are evaluated by using a manual evaluation carried out by 6 experts
in the field in the case of technical specification books.

Knowledge
extraction

Analysis of texts dedicated to children :   A new research activity in the team has
conducted in 2018 about language and children. First, within a collaboration with the
newspaper Le P'tit Libé, we have investigated the use of temporal adverbial phrases in
news articles dedicated to 7-12 years old children. Then, a parallel work studied the
prediction of age recommendations related to texts in order to prevent children to read
texts that they cannot understand due to their intellectual development.

Knowledge
extraction,
multi-level
representa-
tion

Supervised and unsupervised machine learning for intrusion detection [21]:
Data mining techniques play an increasing role in the intrusion detection by analyzing
network data and classifying it as 'normal' or 'intrusion'. In recent years, several data
mining techniques such as supervised, semi-supervised and unsupervised learning are
widely used to enhance the intrusion detection. In this work we have proposed a hybrid
intrusion detection (kM-RF) which outperforms in overall, according to our experimen-
tation, the alternative methods through the accuracy, detection rate and false alarm
rate. A benchmark intrusion detection dataset (ISCX) used to evaluate the efficiency
of the kM-RF, and a deep analysis is conducted to study the impact of the importance
of each feature defined in the pre-processing step.

Knowledge
extraction

**Analogical and Relational proportions between objects or/and attributes
[5, 6]:**   Analogical proportions are statements of the form "A is to B as C is to D",
where A, B, C, D are items of the same nature, or not. In collaboration with Laurent
Miclet, former Emeritus Professor at University of Rennes 1, and Henri Prade , CNRS
Research Director at IRIT, we more particularly consider "relational proportions" of the
form "object A has the same relationship with attribute a as object B with attribute

Knowledge
extraction

b". The nature of relational proportions suggests to handle them in the setting of formal concept analysis. This leads us to the question of defining analogical proportions between formal concepts and studying the links between analogical proportions between formal concepts and analogical proportions between objects or attributes. We derive how relational proportions can be obtained in a formal context from the identification of an analogical complex.
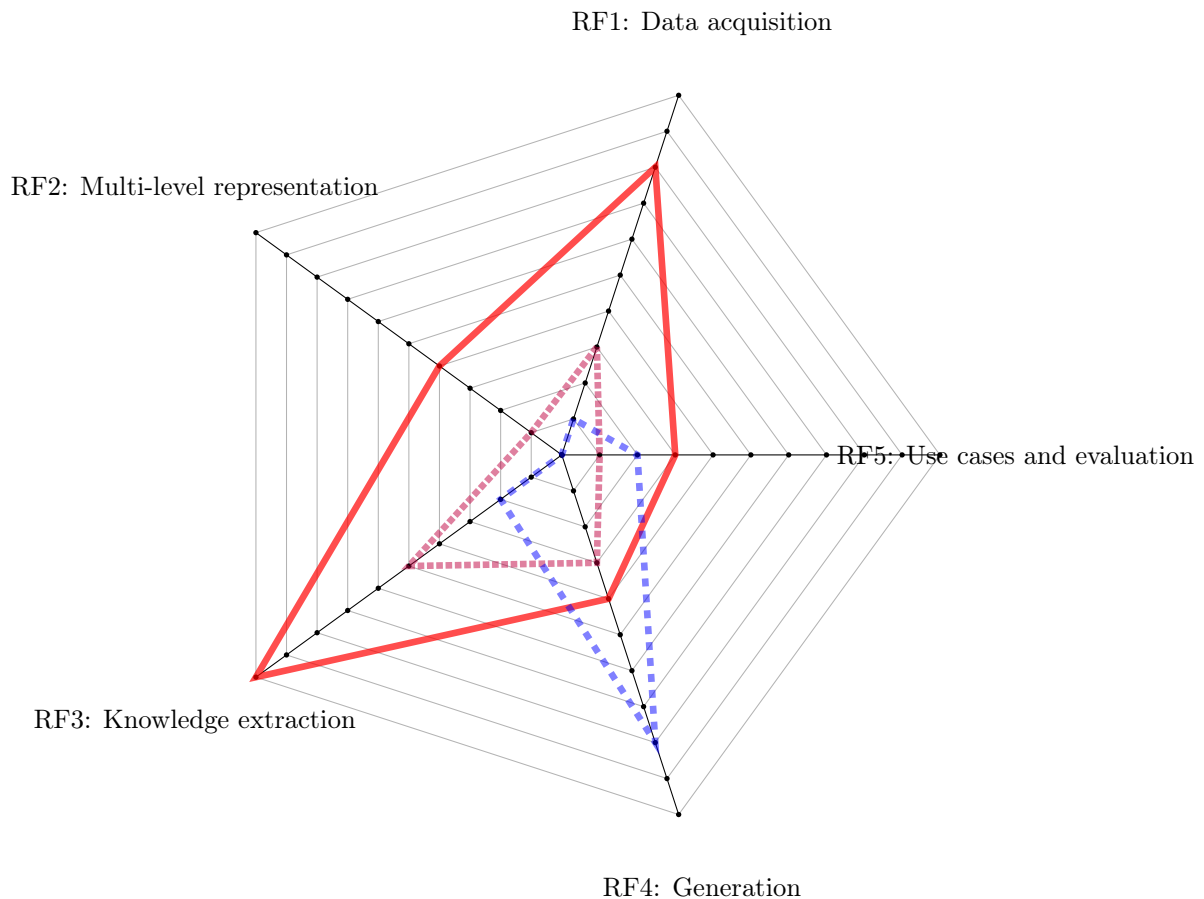
Summary of the contributions

RF1: Data acquisition

RF2: Multi-level representation

RF5: Use cases and evaluation

RF3: Knowledge extraction

RF4: Generation

Figure 2: Contributions to each research focus of the team in 2018 (red, solid) compared to 2017 (purple, dotted) and 2016 (blue, dashed)

3.2 Defended PhDs and HDRs

- Cédric Fayet has defended his PhD on the 18th of December 2018.

3.3   On going PhDs

1. Oussama Ahmia has finished its 2 and a half PhD research program (in the context of a CIFRE grant with the Jurismarchés, renamed OctopusMind, company). His main research contribution this year has been mainly two folds: i) to produce a multi-lingual (partially parallel) corpora consisting of European call for tenders collected from the European TED[2], ii) to proposed some efficient embedding models to categorize (classification and clutering) automatically call for tenders publications. A mixture of Latent Semantic Indexing and Word2vect embedding has been proposed and evaluated with success through Deep Neural Networks (CNN) architecture with attention.These results have been published in LREC2018 and EGC 2019 (long paper).

2. Nicolas Bloyet has completed his 2nd year of research (in the context of a CIFRE grant with the SEED company) addressing the research field of QSAR (quantitative activity - structure relation) that seeks to establish links between observations of the structural nature of a molecule and an "activity", i.e. a physico-chimic property. During this first year, Nicolas Bloyet has finalized the design of a very efficient algorithm for labelled graph fragmentation. Pursuing this similarity with NLP, at the instar of the word2vec approach, a fragment embeddings framework has been set-up to improve regression models for QSAR property prediction. Fisrt experimentation published in EGC2019 (short paper), are very encouraging.

3. Clément Reverdy has completed his last year of PhD. His research addresses the problem of facial expression analysis and synthesis in the context of the animation of LSF signing avatars. He has completed his work on recognition and automatic segmentation of affective facial expression, using (i) classical methods for segmented expressions, and (ii) deep learning methods for non-segmented expressions. He has described his segmentation and annotation models, and discussed his results in [15], using a closed vocabulary of seven expressions.

4. Lucie Naert has completed her second year of PhD. She has developed a whole mocap corpus, called *LSF-Animal*, composed of a set of sentences describing the different properties of animals (location, physical description, behavior, etc.) as well as hand configurations and words composed of several hand configurations. She also developed methods for automatically annotating thirty-two hand configurations, using the *Sign3D* LSF corpus, that she applied to the *LSF-Animal* corpus. She has described her segmentation and annotation models, and discussed her results in [15].

5. Stefania Pecore has completed 2nd year of PhD. The topic of her research is sentiment analysis and, more precisely, detection of opinion from review extracted from French websites. A Movie and Book Annotations corpora in French Language for Aspect Based Sentiment Analysis has been developed and presented at LREC 2018 [16]. Some experiments using classical statistical tools (SVM and Logistic Regression) have suggested directions to follow in order to address the shortcomings of the bag-of-words approach. Some experiment have been conducted around

---

[2]https://ted.europa.eu/TED/main/HomePage.do

the contribution of the negation in opinion mining and the extraction of words and patterns from manually annotated data to enrich a French opinion lexicon. The PhD defense of Stefania Pecore is scheduled the 28th of January 2019.

6. Cédric Fayet has completed his third year. The topic of his research is the detection of anomaly from facial movements and speech signals of a human being. By "anomaly" we mean the existence of foreign elements to a normal situation in a given context. The study focuses in particular on the joint use of facial and vocal expression parameters to detect abnormal variations of expressivity in speech. This year, as a preliminary approach to multimodal anomaly detection, he has focused on the detection of non professional speakers on radio using the acoustic signal (SSPNET-Personality corpus) with several classification methods (GMM, One class SVM and iForest) working on several types of traits (prosodic or Big Five). This work has lead to 3 publications [FDLM17b,FDLM17a,FDLM17c]. The availability of a corpus for the detection of anomalies remaining a problem, Cédric has begun the construction of a specific one containing audio and video data. He published this work to LREC2018 [8]. On the basis of this corpus, he provided a deep experimentation on multimodal anomaly detection, allowing him to defend with great success his PhD on the 18th of December 2018.

7. Antoine Perquin has started his PhD in October. His research addresses new paradigms of speech synthesis opened by recent advances in neural networks and deep learning. The goal of the PhD is to enable generating flexible speech samples based on heterogeneous and massive data. A key aspect within this work lies in properly describing and representing speech variability without relying on expert knowledge. This differs from related work where models are usually trained to produce speech signals, not descriptions, and they are always trained on carefully limited data. During his first year, Antoine compiled pros and cons of many models from the state-of-art, and studied first solutions to adapt them towards the PhD objectives. He also proposed a bottleneck approach for speech synthesis that we published at SLSP 2018 [17].

8. Meysam Shamsi started his PhD in June 2017 and replaced Sandy Aoun after her resignation. His research addresses the optimisation of recording scripts for the expressive reading of audiobooks. The originality of this work is that the problem is addressed by trying to find the best subset of the books we want to synthesize, that will be used to build a voice, then used to generate the remaining part of the books. This way, the goal is to find the best compromise between the size of what we need to record and the quality of the audiobooks we generate. In

[FDLM17b] C. Fayet, A. Delhay, D. Lolive, P.-F. Marteau, "First Experiments to Detect Anomaly Using Personality Traits vs. Prosodic Features", *in : 19th International Conference on Speech and Computer (SPECOM)*, Hatfield, Hertfordshire, United Kingdom, September 2017, https://hal.inria.fr/hal-01583539.

[FDLM17a] C. Fayet, A. Delhay, D. Lolive, P.-F. Marteau, "Big Five vs. Prosodic Features as Cues to Detect Abnormality in SSPNET-Personality Corpus", *in : Interspeech*, Stockholm, Sweden, August 2017, https://hal.inria.fr/hal-01583510.

[FDLM17c] C. Fayet, A. Delhay, D. Lolive, P.-F. Marteau, "Unsupervised Classification of Speaker Profiles as a Point Anomaly Detection Task", *Proceedings of Machine Learning Research*, *74*, p. 152–163, September 2017, https://hal.inria.fr/hal-01631385.

2018, Meysam worked mainly on the setup of an approach to reduce the size of a recording script. To do so, he worked on state of the art approaches and also proposed the use of deep convolutional neural networks or recurrent networks. Interesting results have been achieved and this work should be submitted in early 2019 to a conference.

9. Aghilas Sini has started his PhD in December 2016. He has completed his first year. His research addresses the characterisation and generation of expressivity in function of speaking styles for audiobook synthesis. This Phd takes places in the context of the ANR project SynPaFlex dealing with prosody modelling and the use of prosodic models in speech synthesis. This thesis is also co-funded by the Labex Empirical Foundations of Linguistics (EFL) and co-directed by Elisabeth Delais-Roussarie (DR, CNRS/LLF). During 2017, he worked on the construction of a large speech corpus containing approximately 80 hours of speech. On this topic, we presented a paper at LREC 2018 [20]. This research required the construction of a tool to annotate discourse genres and has been published to the TALN conference [19]. Finally, he also spent 3 months at CSTR, Edinburgh, Scotland, during an internship under the supervision of Prof. Simon King.

10. Tiago Brizolara has started his PhD in November 2018. His research aims at designing interactive applications for live performances from human motion capture. Gesture or full body captured data is used in real time to control multi-sensory outputs such as 2D or 3D visual simulations and sound simulations. To this end, mapping models from captured data and computed descriptors to the parameters of the simulated systems need to be defined. Thanks to deep learning, we shall be able to adapt the models to each individual user for a unique experience. To validate those models, we will propose a methodology based on perceptual evaluations at all levels of the synthesis pipeline (evaluation of the expressive descriptors, of the simulated systems, of the visual or sound effects and of the user experience).

11. Valentin Durand De Gevigney has started his PhD in October 2018. His research aims at designing Deep Neural Network models for the multimodal detection of abnormal behaviours or anomalies. Here, anomalies, are defined has behaviours that are unexpected considering the context they occur in. This work is a follow-up of the PhD of Cédric Fayet on the detection of anomalies in discourse defended in December 2018.

12. Betty Fabre has started her PhD in October 2018. This CIFRE PhD work is jointly conducted in Orange Labs and IRISA/Expression in Lannion. Her research aims at reinforcement learning and structured prediction applied to paraphrase generation.

13. Jade Mekki has started her PhD in September 2018, in the scope of the ANR project TREMoLo. She is supervised by Nicolas Béchet, Gwénolé Lecorvé, and Delphine Battistelli (from university of Paris-Nanterre, MoDyCo lab). Jade works on the characterization of language registers using sequential pattern mining.

14. Ahmad Alaa Al-Dein started his PhD in January 2016 and is currently enrolled in third year. The original subject was on extracting relationships from textual

resources for ontology learning. His work is currently much more focused on hypernym/hyponym relationships found in textual resources. However, it is planned to adapted the techniques developed for hypernym/hyponym relationship extraction to general relationships. Concerning hypernym/hyponym relationships a first extraction system based on earlier work, has been submitted to SemEval competition [10]. A different system has been presented at the KEOD conference [11] and the paper received the best student award.

4   Software development

Animal Instinct   To reach a larger audience with our LSF motion synthesis library (SGN), and to test our algorithms and validate our results, we have developed a serious game called *Animal Instinct*. So far, we have two levels available (more levels will be added in the coming year). In the first level, the player is shown an avatar performing the sign of an animal in LSF together with three photographs of animals: the player has to pick the correct one. In the second level, it's the opposite. The player is shown one photograph and three avatars performing three different signs. Again, the player has to choose the correct answer.

To this end, we have captured a new corpus called *LSF-Animal* containing signs of various animals as well as LSF descriptions of real photographs of animals. The data has then been post-processed and annotated: manually annotated at the gloss level, and automatically annotated for the hand configurations. It is then added to our motion database. For motion retrieval and synthesis, we use the SGN library that has been developed in the team for several years.

5   Contracts and collaborations

5.1   National Initiatives

5.1.1   SynPaFlex ANR project

Participants: Damien Lolive, Gwénolé Lecorvé, Marie Tahon, Gaëlle Vidal, Aghilas Sini.

Expression is leader of a ANR project named SynPaFlex and accepted in July 2015 and started the 1st of December 2015. This project is targeted at the improvement of Text-To-Speech synthesis engines through two main research axes:

- Pronunciation variants modelling and generation

- Context-adapted prosody modelling and generation

The main targeted applications are in the domains of entertainment (audiobook reading, video games), serious games (virtual environments), language learning (dictation, elocution style) or even for vocal aids designed for handicapped people. This

project is mainly supported by IRISA, coordinated by Damien Lolive and involves members from LLF (Laboratoire de Linguistique Formelle) and from ATILF.

Up-to-date information are available at `https://synpaflex.irisa.fr`.

5.1.2   TREMoLo ANR project

Participants: Gwénolé Lecorvé, Nicolas Béchet, Jonathan Chevelu, Sabiha Tahrat.

EXPRESSION is leader of the ANR project TREMoLo, which has been accepted in December 2016. The project studies the use of language registers and seeks to develop automatic methods towards the transformation of texts from a register to another. To do so, the project proposes to extract linguistic patterns which discriminate a register from another, and to integrate them into a probabilistic automatic paraphrase generation process. The language under study is French.

This project is mainly supported by IRISA, coordinated by Gwénolé Lecorvé and involves a member of MoDyCo (UMR 7114 Modèles, Dynamiques, Corpus), Delphine Battistelli.

5.1.3   VOCAGEN PME project

Participants: Nicolas Béchet, Giuseppe Berio, Rémy Kessler.

This project (funded by Pole Images et Réseaux and Région Bretagne) is focused on the building of a software in the field of the construction, allowing users to automatically fill forms starting for the output of a speech recognition system. To this end a concept and term taxonomy is required covering the construction domain. Expression Team is focused on the development of techniques for automatically extracting a relevant terminology and a list of hypernym/hyponym relationships between terms.

This project is coordinated by ScriptGo compan based in Rennes and involves TyKomz company based in Lannion.

5.1.4   KALIGO Dys

Participants: Gwénolé Lecorvé, Damien Lolive, Arnaud Delhay, Quentin Di-Fant.

This project, coordinated by LearnGO, aims at developing a software usable to detect pupils having dys troubles, like dispraxia, and also focusing on re-mediation. The platform should propose dedicated exercises (work on pressure on the sheet, speed measures, fatigability, etc.). It should also be a practical interface for health specialists enabling the preparation of training sessions, helping to follow and report kids' progress. The second main objective is to develop an ergonomic pen usable for re-education.

5.1.5   PEPS TextToKids

Participants: Gwénolé Lecorvé, Damien Lolive, Nicolas Béchet, Jonathan Chevelu.

The PEPS TextToKids project is dedicated to communication with children. The aim of PEPS is to initiate multidisciplinary teamwork, mainly between IRISA (UMR 6074), MoDyCo (UMR 7114) and LLing (UMR 6310) laboratories. The dynamics initiated will allow more ambitious work in terms of necessary means and scientific risks. In particular, a more distant goal is to open the problem to oral and visual modalities to assist the production of multimedia content for children.

5.2   International Initiatives

5.2.1   H2020 NADINE

Participants: Damien Lolive, Gwénolé Lecorvé, Arnaud Delhay, Waseem Safi.

NADINE project aim is to develop a novel way of integrating migrants and refugees through ICT-enabled solutions that will automatically adapt to the specificities of each person. The consortium agrees that one of the main enablers of migrants/refugees inclusion, in the host societies, is their ability to work. Hence NADINE's motto is "Give migrants and refugees their dignity back by giving them a decent job with a decent salary". Taking into account this important factor, NADINE will create an adaptable platform able to: 1) Provide functionalities for skill assessment, 2) dynamically create tailored suited training programs to adapt existing skills into host societies needed skills, 3) provide a digital companion that will suggest and assist the end-users through administrative tasks and 4) create a data lake available to public administration bodies for better organisation of migration flows. NADINE will innovate in several directions from novel training tools, adaptable to different learning setups, to novel ways of information flow handling for public administrations to work efficiently in both business as usual contexts and migration bursts ones. NADINE platform will create potential new markets in different market areas and also will provide novel open tools that will foster new innovation capacity to the EU area.

5.3   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. We have co-supervised the Master level internship of Fédéric Le Bellour in 2018.

- Since PEPS APA in 2013, Nelly Barbot collaborates with Henri Prade, research director of CNRS at IRIT Toulouse and Laurent Miclet (former) emeritus professor at University of Rennes 1 about the formalization of analogical proportions and relational proportions in the framework of formal concept lattice.

5.4   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 continue this collaboration in 2018 with Sébastien Le Maguer, who has moved to the ADAPT center, Trinity College Dublin.

- We also carried on our collaboration with John D. Kelleher from DIT Dublin / ADAPT research centre. Damien Lolive has been welcomed in May 2018 at DIT to discuss possible collaborations on Language Modelling tools using Deep Neural Networks. We also welcomed John D. Kelleher who visited our team in Lannion in December to set up a PhD agreement starting in early 2019.

- The collaboration with McGill University (IDMIL) continued. In 2018, we were thus able to exploit the captured motion of percussionists' movements registered at IDMIL, and to evaluate the performances of the musicians, both by experts and trained students from McGill.

6   Dissemination

6.1   Involvement in the Scientific Community

- Pierre-François Marteau served as a reviewer in international journals (IEEE Trans. on NNLS, Journal of Soft Computing, Pattern Recognition, IEEE Trans. on IFS). He serves as an expert for French Ministry of Research (CIR/JEI) and ANRT (CIFRE), and fr the HCERES (Evaluation of the MIA laboratory). He was member of a thesis committee at Grenoble Alpes University, INSA Lyon, and participated in a few local PhD defense juries. He is member of the Strategic Orientation Committee at IRISA and member of the scientific committee at Université de Bretagne Sud.

- Sylvie Gibet has served as a reviewer for international conferences, including Motion Computing (MOCO2018), and LREC 2018 (Workshop on Sign Language Ressources). She also served as a reviewer for the Journal of Biomechanics.

- Jonathan Chevelu has served as a reviewer for the International conference of the International Speech Communication Association (Interspeech), the International Conference on Audio, Speech and Signal Processing (ICASSP). He served as an expert for the French research agency (ANR).

- Arnaud Delhay has been re-elected as a member of the 'Commission Recherche' (Research comittee) of the IUT of Lannion in November 2015. He has served as a reviewer for the International conference of the International Speech Communication Association (Interspeech), the International Conference on Audio, Speech and Signal Processing (ICASSP).

- Caroline Larboulette is a member of various program committees for international conferences (ACHI2019), 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, Arts).

She is a member of the ACM SIGGRAPH Specialized Conferences Committee that attributes the ACM SIGGRAPH labels to conferences and supervises the budget of conferences sponsored by ACM SIGGRAPH.

- Gwénolé Lecorvé is an elected member of the laboratory council of IRISA, and of the board of directors of the French speech communication association (AFCP). He also serves as a reviewer conferences and journals (Interspeech, ICASSP, *Traitement Automatique des Langues* journal). In 2018, he was part of the organization committee for the conference CORIA-TALN-RJC 2018, and program committee for JEP 2018.

- Damien Lolive is an elected member of the 'Conseil Scientifique' (Scientific council) of ENSSAT, Lannion, and of board of directors of the French speech communication association (AFCP). He serves as a reviewer for the IEEE Transactions on Speech and Language processing, 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) and for the *Journées d'Études sur la Parole* conference. He has also served has an expert for the french research agency, ANR.

- Nelly Barbot has served as a reviewer for the International conference of the International Speech Communication Association (Interspeech).

6.2   Teaching

- Nicolas Béchet teaches various computer sciences courses at the Statistique et Informatique Décisionnelle department of IUT Vannes.

- 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 and web server-side programming in Licence level at École Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT).

- Sylvie Gibet teaches the following Computer Science courses at the faculty of sciences, Université Bretagne Sud: algorithmic at Bachelor level (Python), bases of digital signal processing (1st year master level), and Movement and Artifical Intelligence (2nd year master level).

- 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.

- 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.

- Jean-François Kamp teaches human-computer interaction, programming at the computer science department of IUT Vannes. He is responsible for student internships.

- Caroline Larboulette teaches character animation as part of an introductory lecture of computer science for freshmen and logic for undergraduates of the UFR SSI; C++ programming for ENSIBS graduate students; introduction to computer graphics (I2G) and simulation and interactive applications (SAI) at the master level (Master of Computer Science, AIDN (Interactive Applications and Data)).

- Gwénolé Lecorvé teaches the following computer science courses at École Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT): distributed algorithmics; artificial intelligence; and machine learning in Master level. He also teaches automatic speech recognition and speech synthesis in Research Master program of University of Rennes 1, in Rennes.

- 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.

- Pierre-François Marteau teaches programming languages, logics, introduction to cryptography and information retrieval, machine learning and intrusion detection courses in computer sciences License and Master levels, mostly at École Nationale Supérieure de Bretagne Sud.

- Gildas Ménier teaches various computer sciences courses at the faculty of sciences, Université de Bretagne Sud.

6.3   Conferences, workshops and meetings, invitations

- Pierre-François Marteau, Nicolas Béchet, and Gwénolé Lecorvé were part of the organization committee of CORIA-TALN-RJC 2018, in Rennes.

- Sylvie Gibet attended the Workshop "Training motor skills in virtual reality", inter-GT Animation and Simulation (GT AS) and Virtual Reality (GT RV), Rennes, April 18-19.

- Caroline Larboulette participated to the workshop J.Enaction 2018 (Journée sur l'enaction en animation, simulation et réalité virtuelle), and co-animated with Gilles Col the session on enactivism and signing avatars, Poitiers, November 16 [12].

6.4   Graduate Student and Student internship

- Frédéric Le Bellour has done his Master level internship in Lannion and in collaboration with the LIUM in Le Mans, France. He has been supervised by Damien Lolive and has worked on the semi-supervised annotation of emotions in audiobooks.

- Henri Lasselin has done his Master level internship in Lannion, supervised by Gwénolé Lecorvé. He has been working on the insertion of disfluencies into a fluent text to be synthesized using sequence-to-sequence neural networks.

7   Bibliography

Major publications by the team in recent years

[1] N. BARBOT, V. BARREAUD, O. BOËFFARD, L. CHARONNAT, A. DELHAY, S. LE MA-GUER, D. LOLIVE, "Towards a Versatile Multi-Layered Description of Speech Corpora Using Algebraic Relations", *in: Proceedings of Interspeech*, p. 1501–1504, 2011.

[2] N. BARBOT, O. BOËFFARD, J. CHEVELU, A. DELHAY, "Large Linguistic Corpus Reduction with SCP algorithms", *Computational Linguistics 41*, 3, 2015, p. 30, `https://hal.inria.fr/hal-01135089`.

[3] O. BOËFFARD, L. CHARONNAT, S. LE MAGUER, D. LOLIVE, "Towards Fully Automatic Annotation of Audio Books for TTS", *in: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*, 2012.

[4] A. BOUËNARD, S. GIBET, M. M. WANDERLEY, "Hybrid Inverse Motion Control for Virtual Characters Interacting with Sound Synthesis - Application to Percussion Motion", *Visual Computer Journal 28*, 4, April 2012, p. 357–370, `https://hal.archives-ouvertes.fr/hal-00763280`.

[5] A. BOUËNARD, M. M. WANDERLEY, S. GIBET, F. MARANDOLA, "Virtual Control and Synthesis of Music Performances: Qualitative Evaluation of Synthesized Timpani Exercises", *Computer Music Journal 35*, 3, September 2011, p. 57–72, `https://hal.archives-ouvertes.fr/hal-00763271`.

[6] A. BOUËNARD, M. M. WANDERLEY, S. GIBET, "Virtual Gesture Control of Sound Synthesis: Analysis and Classification of Percussion Gestures", *Acta Acustica united with Acustica 96*, 4, July 2010, p. 668–677, `https://hal.archives-ouvertes.fr/hal-00763258`.

[7] P. CARRENO, S. GIBET, P.-F. MARTEAU, "Synthèse de mouvements humains par des méthodes basées apprentissage : un état de l'art", *Revue Electronique Francophone d'Informatique Graphique 8*, 1, 2014, p. 1–19, `https://hal.archives-ouvertes.fr/hal-01164457`.

[8] J. CHEVELU, G. LECORVÉ, D. LOLIVE, "ROOTS: a toolkit for easy, fast and consistent processing of large sequential annotated data collections", *in: Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland, 2014, `http://hal.inria.fr/hal-00974628`.

[9] J. CHEVELU, D. LOLIVE, S. LE MAGUER, D. GUENNEC, "How to Compare TTS Systems: A New Subjective Evaluation Methodology Focused on Differences", *in: Interspeech*, Dresden, Germany, September 2015, `https://hal.inria.fr/hal-01199082`.

[10] S. GIBET, P. CARRENO-MEDRANO, P.-F. MARTEAU, "Challenges for the Animation of Expressive Virtual Characters: The Standpoint of Sign Language and Theatrical Gestures", *in: Tracts in Advanced Robotics*, 2015, `https://hal.archives-ouvertes.fr/hal-01367757`.

[11] S. GIBET, N. COURTY, K. DUARTE, T. LE NAOUR, "The SignCom System for Data-driven Animation of Interactive Virtual Signers : Methodology and Evaluation".

[12] S. GIBET, F. LEFEBVRE-ALBARET, L. HAMON, R. BRUN, A. TURKI, "Interactive Editing in French Sign Language Dedicated to Virtual Signers: Requirements and Challenges", *Universal Access in the Information Society 15*, 4, 2016, p. 525–539, `https://hal.archives-ouvertes.fr/hal-01367723`.

[13] S. GIBET, P.-F. MARTEAU, K. DUARTE, "Toward a Motor Theory of Sign Language Perception", *Human-Computer Interaction and Embodied Communication, GW 2011 7206*, 2012, p. 161–172.

[14] D. GUENNEC, D. LOLIVE, "On the suitability of vocalic sandwiches in a corpus-based TTS engine", *in: Interspeech*, San Francisco, United States, September 2016, `https://hal.inria.fr/hal-01338839`.

[15] G. KE, P.-F. MARTEAU, G. MÉNIER, "Improving the clustering or categorization of bi-lingual data by means of comparability mapping", October 2013.

[16] T. LE NAOUR, N. COURTY, S. GIBET, "Cinématique guidée par les distances", *Revue Electronique Francophone d'Informatique Graphique Vol. 6*, N° 1, 2012, p. 15–25, `https://hal.archives-ouvertes.fr/hal-00762919`.

[17] T. LE NAOUR, N. COURTY, S. GIBET, "Spatio-temporal coupling with the 3D+t motion Laplacian", *Computer Animation and Virtual Worlds*, 2013, p. CAV1518, `https://hal.archives-ouvertes.fr/hal-00861779`.

[18] G. LECORVÉ, D. LOLIVE, "Adaptive Statistical Utterance Phonetization for French", *in: Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 5 p., 2 columns, Brisbane, Australia, April 2015, `https://hal.inria.fr/hal-01109757`.

[19] D. LOLIVE, N. BARBOT, O. BOËFFARD, "B-spline model order selection with optimal MDL criterion applied to speech fundamental frequency stylisation", *IEEE Journal of Selected Topics in Signal Processing 4*, 3, 2010, p. 571–581.

[20] P.-F. MARTEAU, S. GIBET, "On Recursive Edit Distance Kernels with Application to Time Series Classification", February 2013.

[21] R. QADER, G. LECORVÉ, D. LOLIVE, P. SÉBILLOT, "Probabilistic Speaker Pronunciation Adaptation for Spontaneous Speech Synthesis Using Linguistic Features", *in: International Conference on Statistical Language and Speech Processing (SLSP)*, p. 12 p., 1 column, Budapest, Hungary, November 2015, `https://hal.inria.fr/hal-01181192`.

Articles in referred journals and book chapters

[1] P. CARRENO MEDRANO, S. GIBET, P.-F. MARTEAU, "Perceptual Validation for the Generation of Expressive Movements from End-Effector Trajectories", *ACM Transactions on Interactive Intelligent Systems (TiiS)*, July 2018, `https://hal.archives-ouvertes.fr/hal-01854494`.

[2] M. TAHON, G. LECORVÉ, D. LOLIVE, "Can we Generate Emotional Pronunciations for Expressive Speech Synthesis?", *IEEE Transactions of Affective Computing*, April 2018, `https://hal.archives-ouvertes.fr/hal-01802463`.

Publications in Conferences and Workshops

[3] O. AHMIA, N. BÉCHET, P.-F. MARTEAU, "Two Multilingual Corpora Extracted from the Tenders Electronic Daily for Machine Learning and Machine Translation Applications", *in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Myazaki, Japan, May 2018, `https://hal.archives-ouvertes.fr/hal-01865091`.

[4] P. ALAIN, G. LECORVÉ, D. LOLIVE, A. PERQUIN, "The IRISA Text-To-Speech System for the Blizzard Challenge 2018", *in : Blizzard Challenge Workshop*, Hyderabad, India, September 2018, `https://hal.inria.fr/hal-01998037`.

[5] N. BARBOT, L. MICLET, H. PRADE, "Relational proportions between objects and attributes", *in : 6th Workshop "What can FCA do for Artificial Intelligence?", IJCAI/ECAI Workshop*, S. O. Kuznetsov, A. Napoli, S. Rudolph (editors), *2149*, p. 33–44, Stockhom, Sweden, July 2018, `https://hal.inria.fr/hal-01985031`.

[6] N. BARBOT, L. MICLET, H. PRADE, "The analogical cube of opposition", *in : SQUARE 2018 : 6th World Congress on the Square of Opposition*, J.-Y. Beziau, A. Buchsbaum, I. Vandoulakis (editors), Crete, Greece, November 2018, `https://hal.inria.fr/hal-01984982`.

[7] L. CHEN, S. GIBET, C. MARTEAU, "CONDUCT: An Expressive Conducting Gesture Dataset for Sound Control", *in : Language Resources and Evaluation Conference (LREC 2018)*, Miyazaki, Japan, May 2018, `https://hal.archives-ouvertes.fr/hal-01851389`.

[8] C. FAYET, A. DELHAY, D. LOLIVE, P.-F. MARTEAU, "EMO&LY (EMOtion and AnomaLY) : A new corpus for anomaly detection in an audiovisual stream with emotional context", *in : Language Resources and Evaluation Conference (LREC)*, Miyazaki, Japan, May 2018, `https://hal.inria.fr/hal-01685055`.

[9] S. GIBET, "Building French Sign Language Motion Capture Corpora for Signing Avatars", *in : Workshop on the Representation and Processing of Sign Languages: Involving the Language Community, LREC 2018*, Miyazaki, Japan, May 2018, `https://hal.archives-ouvertes.fr/hal-01851398`.

[10] A. ISSA, M. HARZALLAH, G. BERIO, N. BÉCHET, A. FAOUR, "A EXPR : A Combined Approach for Hypernym Discovery", *in : The 12th International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, United States, June 2018, `https://hal.archives-ouvertes.fr/hal-01994866`.

[11] A. ISSA, M. HARZALLAH, G. BERIO, N. BÉCHET, A. FAOUR, "Redefining Hearst Patterns by using Dependency Relations", *in : 10th International Conference on Knowledge Engineering and Ontology Development*, Seville, France, September 2018, `https://hal.archives-ouvertes.fr/hal-01920043`.

[12] C. LARBOULETTE, S. GIBET, "Avatar signeurs : que peuvent-ils nous apprendre ?", *in : J.Enaction 2018 : Journée sur l'énaction en animation, simulation et réalité virtuelle*, Poitiers, France, November 2018, `https://hal.archives-ouvertes.fr/hal-01990023`.

[13] G. LECORVÉ, H. AYATS, B. FOURNIER, J. MEKKI, J. CHEVELU, D. BATTISTELLI, N. BÉCHET, "Construction conjointe dâun corpus et dâun classifieur pour les registres de langue en français", *in : Traitement automatique du langage naturel (TALN)*, Rennes, France, May 2018, `https://hal.inria.fr/hal-02002601`.

[14] J. MEKKI, D. BATTISTELLI, G. LECORVÉ, N. BÉCHET, "Feature identification for register characterization", *in : Rencontre des jeunes chercheurs en traitement automatique du langage naturel et recherche d'information (CORIA-TALN-RJC)*, Rennes, France, May 2018, `https://hal.inria.fr/hal-02002612`.

[15] L. NAERT, C. REVERDY, C. LARBOULETTE, S. GIBET, "Per Channel Automatic Annotation of Sign Language Motion Capture Data", *in : Workshop on the Representation and Processing of Sign Languages: Involving the Language Community, LREC 2018*, Miyazaki, Japan, May 2018, `https://hal.archives-ouvertes.fr/hal-01851404`.

[16] S. PECORE, J. VILLANEAU, "Complex and Precise Movie and Book Annotations in French Language for Aspect Based Sentiment Analysis", *in : LREC 2018*, MIYAZAKI, Japan, May 2018, `https://hal.archives-ouvertes.fr/hal-01968572`.

[17] A. PERQUIN, G. LECORVÉ, D. LOLIVE, L. AMSALEG, "Phone-Level Embeddings for Unit Selection Speech Synthesis", *in : SLSP 2018 - 6th International Conference on Statistical Language and Speech Processing*, Springer, p. 21–31, Mons, Belgium, October 2018, `https://hal.archives-ouvertes.fr/hal-01840812`.

[18] R. QADER, G. LECORVÉ, D. LOLIVE, P. SÉBILLOT, "Disfluency Insertion for Spontaneous TTS: Formalization and Proof of Concept", *in : SLSP 2018 - 6th International Conference on Statistical Language and Speech Processing*, Springer, p. 1–12, Mons, Belgium, October 2018, `https://hal.inria.fr/hal-01840798`.

[19] A. SINI, E. DELAIS-ROUSSARIE, D. LOLIVE, "Annotation automatique des types de discours dans des livres audio en vue d'une oralisation par un système de synthèse", *in : TALN-RECITAL 2018* , Rennes, France, May 2018, `https://hal.archives-ouvertes.fr/hal-01848856`.

[20] A. SINI, D. LOLIVE, G. VIDAL, M. TAHON, E. DELAIS-ROUSSARIE, "SynPaFlex-Corpus: An Expressive French Audiobooks Corpus Dedicated to Expressive Speech Synthesis", *in : Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018, `https://hal.archives-ouvertes.fr/hal-01826690`.

[21] S. SOHEILY-KHAH, P.-F. MARTEAU, N. BÉCHET, "Intrusion Detection in Network Systems Through Hybrid Supervised and Unsupervised Machine Learning Process: A Case Study on the ISCX Dataset", *in : 2018 1st International Conference on Data Intelligence and Security (ICDIS)*, IEEE, South Padre Island, France, April 2018, `https://hal.archives-ouvertes.fr/hal-01803772`.

[22] M. TAHON, D. LOLIVE, "Discourse phrases classification: direct vs. narrative audio speech", *in : Speech Prosody*, Poznan, Poland, June 2018, `https://hal.archives-ouvertes.fr/hal-01790910`.

Internal Reports

[23] H. LASSELIN, G. LECORVÉ, "Make text look like speech: disfluency generation using sequence-to-sequence neural networks", *Research report*, Univ Rennes, CNRS, IRISA, France ; IRISA, équipe EXPRESSION, January 2018, `https://hal.inria.fr/hal-01738344`.

Miscellaneous

[24] H. LASSELIN, G. LECORVÉ, *Make text look like speech: disfluency generation using sequence-to-sequence neural networks Domain: Document and Text Processing-Machine Learning*, Mémoire, Université de Rennes 1 [UR1], August 2018, `https://hal.inria.fr/hal-02002541`.