UMR IRISA

Team EXPRESSION

Expressiveness in Human Centered Data/Media

Vannes, Lannion & Lorient

Activity Report

2015

Contents

1 Team

1.1 Composition

Head of the team

Pierre-François Marteau, Professor, Université de Bretagne Sud

Administrative assistant

Sylviane Boisadan, Université de Bretagne Sud

Permanent members

Nelly Barbot, Associate professor, Université de Rennes 1

Nicolas Béchet, Associate professor, Université de Bretagne Sud

Giuseppe Bério, Professor, Université de Bretagne Sud

Jonathan Chevelu, Associate professor, Université de Rennes 1 (since September 2015)

Arnaud Delhay-Lorrain, Associate professor, Université de Rennes 1

Sylvie Gibet, Professor, Université de Bretagne Sud

Caroline Larboulette, Associate professor, Université de Bretagne Sud (since September 2015)

Gwénolé Lecorvé, Associate professor, Université de Rennes 1

Damien Lolive, Associate professor, Université de Rennes 1

Gildas Ménier, Associate professor, Université de Bretagne Sud

Jeanne Villaneau, Associate professor, Université de Bretagne Sud

Associate members

Vincent Barreaud, Associate professor, Université de Rennes 1

Elisabeth Delais-Roussarie, Senior researcher, CNRS/LLF (since June 2015)

Jean-François Kamp, Associate professor, Université de Bretagne Sud

Farida Said, Associate professor, Université de Bretagne Sud

Non-permanent members

Jonathan Chevelu, ATER, Université de Rennes 1 (until August 2015)

Nehla Ghouaiel, ATER, Université de Bretagne Sud

Ioana Suciu, ATER, Université de Rennes 1 (until August 2015)

Marie Tahon, Post-doctoral researcher, Université de Rennes 1 (since December 2015)

Gaëlle Vidal, Engineer, Université de Rennes 1 (since December 2015)

PhD candidates

Yonatan Carranza Alarcón, Université de Bretagne Sud, 1st year

Pamela Carreno, Université de Bretagne Sud, ANR INGREDIBLE, 3rd year

Lei Chen, Université de Bretagne Sud/Univ. McGill, ARED, 3rd year

Marc Dupont, Université de Bretagne Sud, Thèse CIFRE Thales, 2nd year

Cédric Fayet, Université de Rennes 1, 1st year

David Guennec, Université Rennes 1, 4th year

Stefania Pecóre, Université de Bretagne Sud, 1st year

Raheel Qader, Université de Rennes 1, 3rd year

Clément Reverdy, Université de Bretagne Sud, CDE+Labo, 2nd year

Hai Hieu Vu, Université de Bretagne Sud, 5th year

Master students
> *Inès Dabbebi, Université de Bretagne Sud*
> *Cédric Fayet, Université de Rennes 1*

1.2 Evolution of the staff

The permanent staff has positively evolved in 2015 with the hiring of two associate professors: Caroline Larboulette has joined the Gesture research activity at Université Bretagne Sud (UFR SSI Vannes), while Jonathan Chevellu has joined the Speech research activity at Université de Rennes1 (ENSSAT Lannion). Regarding the associate members, Nehla Ghouaiel, has been hired ATER at the Université Bretagne Sud (Vannes) in September 2014 and joined EXPRESSION since then on gesture recognition activity. Ioana Suciu, has been hired ATER at Université de Rennes 1 and joined EXPRESSION. The number of PhD students is slightly increasing : Cédric Fayet (Université de Rennes 1 (Lannion)), Stefania Pecóre (Université Bretagne Sud (Vannes)) have been hired as PhD students at IRISA, in October 2015.

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 proposed 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.

3 New Results

3.1 Main events

2015 has been the year of the labelisation by IRISA of the Expression Team Project [1].

---

[1] https://www-expression.irisa.fr/files/2011/07/expression_team_submitted.pdf

3.2   New Results by Key Issues

In accordance with the Team Project, the main outcomes for 2015 are listed into the following key issues items defined for the team:

Data acquisition:

1. Hand gesture: Using the VHAND data glove, two new experimental data set have been built in the scope of Marc dupont's PhD. The first data set contains "static gestures" (produced statically, namely without mobility): it contains 1 subject, 4 data streams, each stream includes 6 repetitions of each gesture (among a vocabulary of 8 gestures). The second data set contains have been produced statically and in mobility (while walking and running): it contains 3 subjects 6 flows (2 Static, 2 walks, 2 runs), a scenario with several repetitions (at least 2) of each gesture.

2. Multi-subjects mocap recording: the two-subjects database containing expressive theatrical gestures has been extended. Two main changes in the recording protocol have been introduced:

   - In addition to the three sequences already defined in the magician scenario, we decided to include one walk motion at least one minute long and a short improvisation sketch freely chosen by each actor. The former will allow us to compare our results to previous work, since locomotion, in particular walking, has been extensively studied in both perception and automatic recognition of affect/emotions. The latter sequence will provide us with measurements about the end-effector trajectories behavior on considerably different movement behaviors. Combined, we can use them to evaluate the expressiveness of the actors and the benefits of using motions inspired by theatrical scenarios.

   - Five additional actors (two females and three males, ranging in age from 38 to 54) were asked to perform the five motion sequences already defined. Once again, actors were asked to convey both emotions and meaning solely through body motions. To facilitate the enacting of all emotions during the sequences, a combined induced mood procedure was defined. We used a story-based and imagination-based mood induction procedures.

3. Large Linguistic Corpus Reduction with SCP algorithms: Linguistic corpus design is a critical concern for building rich annotated corpora useful in different domains of applications. For example, speech technologies such as ASR (Automatic Speech Recognition) or TTS (Text-to-Speech) need a huge amount of speech data to train data-driven models or to produce synthetic speech. Collecting data is always related to costs (recording speech, verifying annotations, etc.), and as a rule of thumb, the more data you gather, the more costly your application will be. Within this context, in [7], we present solutions to reduce the amount of linguistic text content while maintaining a sufficient level of linguistic richness required by a model or an application. This problem can be formalized as a Set Covering Problem (SCP) and we evaluate two algorithmic heuristics applied to design large text corpora in English and French for covering phonological information or POS labels. The first considered algorithm is a standard greedy solution with an agglomerative/spitting strategy and we propose a second algorithm based on Lagrangian relaxation. The latter approach provides a lower bound to the cost of each covering solution. This lower bound can be used as a metric to evaluate the quality of a reduced corpus whatever the algorithm applied. Experiments show that a sub-optimal algorithm like a greedy algorithm achieves good results; the cost of its solutions is not so far from the lower bound (about 4.35% for 3-phoneme coverings). Usually, constraints in SCP are binary; we proposed here a generalization where the constraints on each covering feature can be multi-valued.

Multi-level representations:

1. Motion descriptors: In 2015, we have done a review of computable descriptors of human motion. We have classified them into two categories: low-level descriptors that compute quantities directly

from the raw motion data ; and high-level descriptors that use low-level ones to compute boolean, single value or continuous quantities that can be interpreted, automatically or manually, to qualify the meaning, style or expressiveness of a motion. We provide formulas inspired from the state of the art that can be applied to 3D motion capture data.

Data mining and knowledge extraction

1. Exploiting Sparsity in Gesture Time Series (Coarse-DTW): Dynamic Time Warping (DTW) is considered as a robust measure to compare numerical time series when some time elasticity is required. Even though its initial formulation can be slow, extensive re-search has been conducted to speed up the calculations. However, those optimizations are not always available for multidimensional time series. In this paper, we focus on time series describing gesture movement, all of which are multidimensional. Our approch propose to speed up the processing by 1. adaptively downsampling the time series into sparse time series and 2. generalizing DTW into a version exploiting sparsity. Furthermore, the downsampling algorithm doesn't need to know the whole timeseries to function, making it a good candidate for streaming applications such as real-time gesture recognition.

2. We address the problem of mining sequential patterns under multiple constraints. Unlike classical algorithms, our approach handles various types of constraints which are not only numeric but also symbolic and syntactic. These multiple constraints enable us to express a large scope of knowledge to focus on interesting patterns. We illustrate our approach with the detection of gene–rare disease relationships from biomedical texts for the documentation of rare diseases.

3. Digital Humanities make more and more structured and richly annotated corpora available. Most of this data rely on well known and established standards, such as TEI, which especially enable scientists to edit and publish their work. However, one of the remaining problems is to give adequate access to this rich data, in order to produce higher-order knowledge. In this paper, we present an integrated environment combining an advanced search engine and text-mining techniques for hermeneutics in Digital Humanities. Relying on semantic web technologies, the search engine uses full text as well as complex embedding structures and offers a single interface to access rich and heterogeneous data and meta-data. Text-mining possibilities enable scholars to exhibit regularities in corpora. Results obtained on the Cartesian corpus illustrate these principles and tools.

4. Semantic similarity between sentences: Mesuring similarity between short texts (such as sentences) has many interesting applications such as multi-document summarization,conversational agent, etc. We proposed a semantic similarity measure for sentence comparison based on the exploitation of Wikipedia as the only language resource. Such similarity measure is therefore usable for a wide range of languages, basically those covered by Wikipedia. Random Indexing is used to cope with the great dimensionality and the spareness of the data vectorial representations. Furthermore, a statistical weight function is used to reduce the noise generated by the use of a multi domain corpus such as Wikipedia. This semantic similarity measure has been evaluated on SemEval 2014 dataset for English language leading to very promising results, basically above the average level of the competing systems that exploit Wikipédia in conjunction with other sources of semantic information. It has been also evaluated on a set of pairs of sentences in French that we have build specifically for the task.

Generation:

1. A probabilistic framework for pronunciation generation: Traditional utterance phonetization methods concatenate pronunciations of uncontextualized constituent words. This approach is

too weak for some languages, like French, where transitions between words imply pronunciation modifications. Moreover, it makes it difficult to consider global pronunciation strategies, for instance to model a specific speaker or a specific accent. To overcome these problems, we have proposed a new original phonetization approach for French to generate pronunciation variants of utterances [18]. This approach offers a statistical and highly adaptive framework by relying on conditional random fields and weighted finite state transducers. The approach is evaluated on a corpus of isolated words and a corpus of spoken utterances.

2. Probabilistic pronunciation modeling for spontaneous speech: Pronunciation adaptation consists in predicting pronunciation variants of words and utterances based on their standard pronunciation and a target style. This is a key issue in text-to-speech as those variants bring expressiveness to synthetic speech, especially when considering a spontaneous style. We have proposed a new pronunciation adaptation method which adapts standard pronunciations to the style of individual speakers in a context of spontaneous speech [20]. Its originality and strength are to solely rely on linguistic features and to consider a probabilistic machine learning framework, namely conditional random fields, to produce the adapted pronunciations. Features are first selected in a series of experiments, then combined to produce the final adaptation method. Backend experiments on the Buckeye conversational English speech corpus show that adapted pronunciations significantly better reflect spontaneous speech than standard ones, and that even better could be achieved if considering alternative predictions.

3. Adaptive duration target cost for TTS: Unit selection speech synthesis systems generally rely on target and concatenation costs for selecting a best unit sequence. These costs, though often considering contextual features, mainly include local distances that are accumulated afterwards. In [15], we describe a new duration target cost that takes a whole sequence into account. It aims at selecting a sequence globally good, instead of a very good sequence almost everywhere but having a few local duration cost leaps that are counter-balanced by other units. The problem of weighting this new duration cost with other sub-costs is also investigated. Experiments showed this new measure performed well on sentences featuring duration artefacts, while not deteriorating others.

4. Comparing rythmic patterns between read and synthesized speech: In [9], compares the rhythmic features obtained in natural and in synthesized speech along three dimensions : the speech type (synthesized vs natural speech), the literary genre (rhymes, poems vs story telling), and the communication setting (speech addressed to children vs addressed to adults). This study is based on the analysis of duration patterns observed in a corpus recorded by two speakers and generated by a synthesis system. Results show that rhythmic differences between synthesized and natural speech are mostly due to the marking of prosodic structure by duration. This work is achieved within a project aiming at improving the naturalness of synthesized speech for use in pedagogical software.

5. On the effectiveness of corpus reduction for TTS: TTS voice building generally relies on a script extracted from a big text corpus while optimizing the coverage of linguistic and phonological events supposedly related to voice acoustic quality. Previous works have shown differences on objective measures between smartly reduced and random corpora, but not when subjective evaluations are performed. For us, those results do not come from corpus reduction utility but from evaluations that smooth differences. In [12], we highlight those differences in a subjective test, by clustering test corpora according to a distance between signals so as to focus on different synthesized stimuli. The results show that covering appropriate features has a real impact on the perceived quality.

Use cases and evaluation:

1. Speech synthesis system evaluation: Subjective evaluation is a crucial problem in the speech processing community and especially for the speech synthesis field, no matter what system is used. Indeed, when trying to assess the effectiveness of a proposed method, researchers usually conduct subjective evaluations by randomly choosing a small set of samples, from the same domain, taken from a baseline system and the proposed one. When selecting them randomly, statistically, samples with almost no differences are evaluated and the global measure is smoothed which may lead to judge the improvement not significant. To solve this methodological flaw, in [11], we propose to compare speech synthesis systems on thousands of generated samples from various domains and to focus subjective evaluations on the most relevant ones by computing a normalized alignment cost between sample pairs. This process has been successfully applied both in the HTS statistical framework and in the corpus- based approach. We have conducted two perceptive experiments by generating more than 27000 samples for each system under comparison. A comparison between tests involving most different samples and randomly chosen samples shows clearly that the proposed approach reveals significant differences between the systems.

2. International Blizzard challenge: For our first participation to the Blizzard Challenge, we chose to take part in the first task by building one voice for each of the 5 Indian languages out of the 6 requested (Bengali, Hindi, Malayalam, Tamil and Telugu). The process followed to build the voices from given data and the architecture of our system is described in [10] . In particular, we introduce a penalty in the concatenation cost, inherited from the field of corpus covering, in order to block some concatenations based on their phonological class. Moreover, a fuzzy function is proposed to relax the penalty based on the concatenation quality with respect to the cost distribution. Considering that no language specific processing has been done and that we had never worked with Indian languages, results are very satisfying.

3.3   Defended PhDs

Hai Hieu Vu defended his PhD: *Random Indexing and Sentence Similarity applied to automatic multi-document summarization.*

Abstract : With the growing mass of textual data on the Web, automatic summarization of topic-oriented collections of documents has become an important research field of Natural Language Processing. The experiments described in the thesis were framed within this context. Evaluating the semantic similarity between sentences is central to this work and the approach is based on distributional similarity and vector representation of terms, with Wikipedia as a reference corpus. Several similarity measures are proposed, which were evaluated and compared on different data sets: the SemEval 2014 challenge corpus for the English language and built datasets for French. The good performance showed by the measures led to use them in a multi-document summary task, which implements a pagerank-type algorithm. The system was evaluated on the DUC 2007 datasets for English and RPM2 corpus for French. This simple approach, based on a resource readily available in many languages, proved efficient, robust and the encouraging outcomes open up real prospects of improvement.

3.4   On going PhDs

1. Marc Dupont has completed his 2nd year of research addressing gesture recognition using a data glove in the scope of controlling a moving robot in an adverse environment. His main achievements are:

   - the design of two "benchmarks" software for the comparison of hand gesture recognition algorithms: the first one is used to test isolated gestures, while the second one is used for gesture spotting and recognition in stream.

- an improved data base containing isolated gestures and gestures in sequence (data stream).
- a "sparse" approximation of DTW (dynamic time warping) algorithm, called Coarse-DTW, allowing alignment speed-up via a technique of adaptive sub-sampling of the time series.

.

2. David Guennec has completed his third year of research addressing the improvement of cost functions for unit selection speech synthesis. His main achievements concern the use of an A* algorithm and its comparison to the classical Viterbi algorithm and the introduction of a a new target cost to constrain the duration of phonemes. For this last point, phoneme durations are predicted using a neural network and used to constrain the selection process. During this year, his visited the Idiap Research Institute for 5 months. During his stay he worked on prosody modeling with an atom-based model and its integration into the speech synthesis process. He is now ATER at Enssat, University of Rennes 1. His PhD defense is planned to take place mid-2016.

3. Raheel Qader has completed his second year of research addressing phonology modeling for expressive speech synthesis. During his first year, his main achievements were a review of the state of the art, the analysis of a spontaneous speech English corpus and first experiments towards a pronunciation variant predictive model for speaker and style adaptation. This year, he has published an article to the SLSP conference on spontaneous pronunciation prediction. For 2016, we plan to work on the subjective evaluation of pronunciation modifications and also on disfluencies in spontaneous speech.

4. Pamela Carreno has completed her third year of research addressing expressive gesture synthesis using Inverse Kinematics and expressive gesture recognition using LDA and SVM. Based on findings reported by the psychology research community and quantitative results obtained in the computer animation domain during the last years, she has proposed to represent affective bodily movement through a low-dimensional parameterization consisting of the spatio-temporal trajectories of eight main joints in the human body (hands, head, feet, elbows and pelvis). Using a combined evaluation protocol, we has shown that this low-dimensional parameterization and the features derived from it (first, second and third derivatives of position) are a compact and sufficient representation of affective motions that can be used for automatic recognition of affect and the generation of new affective-expressive motions. [?]. She also designed the protocol and recorded the extended theatrical gesture database.

5. Lei Chen has completed her third year of research addressing the analysis and classification of expressive timpani gestures. She has characterized timpani gestures by temporal kinematic features, containing most information responsible for the sound-producing actions. In order to evaluate the feature sets, a classification approach has been conducted under three main attack categories (legato, accent and vertical accent) and sub-categories (dynamics, striking position). Two studies are carried out: intra-subject and inter-subjects classification. Results are presented in terms of a quantitative ranking of students, using professional gestures as training set, and their gestures as test set.

6. Clément Reverdy has started his second year of PhD. His research addresses the problem of facial expression synthesis in the context of sign language. He worked on blendshape-based animation techniques using motion captured data [RGL15a]. In particular, he developed a blendshape-based

[?]   *** ERROR: citation 'carrenomedrano:hal-01367814' undefined ***

[RGL15a]   C. Reverdy, S. Gibet, C. Larboulette, "Animation faciale basée données : un état de l'art", *in:* *28èmes journées de l'Association Française en Informatique Graphique*, *Proceedings of AFIG*, Lyon, France, November 2015, https://hal.archives-ouvertes.fr/hal-01366447.

motion synthesis pipeline and worked on the optimal placement of markers by comparing state-of-the-art layouts, empirical layouts, and automatically designed layouts by implementing a K-means clustering technique [RGL15b].

7. Cédric Fayet has started his PhD. The topic of his research is the detection of abnormality from facial movements and speech signals of a human being. By "abnormality" we mean the existence of foreign elements to a normal situation in a given context. The study will focus in particular on the joint use of facial and vocal expression parameters to detect abnormal variations of expressivity in speech, not only related to emotion, but also to social interactions and psychological signals. These abnormal signals can appear in extreme stress situations for pilots or vehicle drivers, for example. This study could also find applications in the medical field, e.g., detection of abnormal behaviors due to mental disabilities such as autism. We aimed at developing a system capable of detecting abnormal behaviors by the analysis of records of concrete situations.

8. Stefania Pecòre has started his PhD. The topic of his research is the detection of opinion in social networks such as Allocine, Tripadvisor, etc.

4   Software

4.1   SGN

Participants: Caroline Larboulette, Sylvie Gibet.

The SGN library has been enriched with $3D$ rendering of the avatar: a skinning algorithm and a blendshape algorithm.

In addition, a dynamics module has been added for mass-spring and particle simulations to use SGN in a more general context (physically based simulation from captured data, not only avatars).

A QT interface has been built for a signing avatar demonstration based on SGN and interactive contruction of sentences from a sequence of gloses chosen by the user. The demo has first been presented at the public open house of IRISA and later refined and used for the HCERES evaluation of the laboratory.

4.2   ROOTS

Participants: Nelly Barbot, Vincent Barreaud, Jonathan Chevelu, Arnaud Delhay, Sébastien Le Maguer, Gwénolé Lecorvé, Damien Lolive.

The development of new methods for given speech and natural language processing tasks usually faces, beyond scientific aspects, various technical and practical data management problems. Indeed, the sets of required annotated features and their desired distribution in the training data are rarely the same for two different tasks, and many dedicated systems or expert resources use different file formats, time scales, or alphabets of tags.

In this context, ROOTS, stemming for Rich Object Oriented Transcription System, is an open source toolkit dedicated to annotated sequential data generation, management and processing, especially in the field of speech and language processing. It works as a consistent middleware between dedicated data processing or annotation tools by offering a consistent view of various annotation levels and synchronizing them. Doing so, ROOTS ensures a clear separation between description and treatment.

[RGL15b]   C. REVERDY, S. GIBET, C. LARBOULETTE, "Optimal Marker Set for Motion Capture of Dynamical Facial Expressions", *in: MIG2015 8th ACM SIGGRAPH Conference on Motion in Games , Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*, p. 31–36, Paris, France, November 2015, https://hal.archives-ouvertes.fr/hal-01279483.
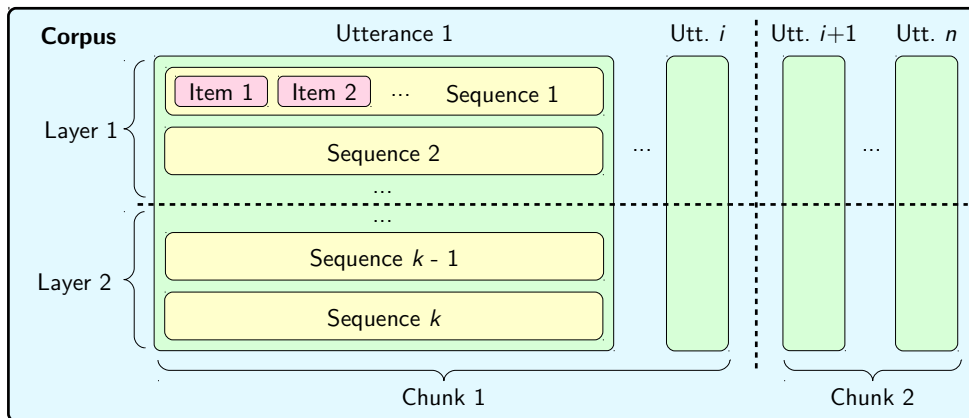
Figure 1: Hierarchical organization of data in ROOTS.

Theoretical aspects of multilevel annotation synchronization have previously been published in [BBB+11] while a prototype had been presented and applied to an audiobook annotation task in [BCLML12].

As summarized in Figure 1, data are organized hierarchically in Roots, starting from fine grain information in items and moving to macroscopic representations as corpora. As a fundamental concept, data in ROOTS is modeled as sequences of items. These items can be of many types, e.g., words, graphemes, named entity classes, signal segments, etc., and can thus represent various annotation levels of the same data. Correspondences between items from different sequences are then defined as algebraic relations, leading to a graph where nodes are items and edges are derived from relations. Then, interrelated sequences are gathered into utterances. According to the application domain, utterances can refer to sentences, breath groups, or any relevant unit. A part of the recent work on ROOTS has focused on extending this hierarchization of data to easily handle large collections of data. Hence, the notion of corpus has been defined as a list of utterances or, recursively, as a list of subcorpora (called chunks), for instance to represent a chapter as a list of paragraphs. Besides chunks, corpora can also be partitioned "horizontally" into layers which gather annotations from a same field. The following operations are allowed for each data hierarchization level:

- Item: get/set the content/characteristics; get other items in relation; dump[2].

- Sequence: add/remove/get/update items; dump,

- Relation: get items related to another; link or unlink items; dump,

- Utterance: add/remove/get/update sequences; add/remove/get/update direct or composed relations; dump,

- Corpus: add/remove/get/update an utterance; add/remove chunks/layers; load/save; dump.

---

[2]Dump refers to input/output operations in raw text, XML and JSON formats.

[BBB+11]   N. Barbot, V. Barreaud, O. Boëffard, L. Charonnat, A. Delhay, S. Le Maguer, D. Lolive, "Towards a Versatile Multi-Layered Description of Speech Corpora Using Algebraic Relations", *in: Conference of the International Speech Communication Association (Interspeech)*, p. 1501–1504, Florence, Italie, 2011.

[BCLML12]   O. Boëffard, L. Charonnat, S. Le Maguer, D. Lolive, "Towards Fully Automatic Annotation of Audio Books for TTS", *in: Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC)*, European Language Resources Association (ELRA), Istanbul, Turkey, may 2012.

Roots is made of a core library and of a collection of utility scripts. All functionalities are accessible through a rich API either in C++ or in Perl. Recently, this API has greatly evolved and to ease building Roots corpora based on this API (e.g., with the notion of corpus), and accessing information in flexible and intuitive manners. Extra developments have also led to the following improvements: new wrapping scripts for basic corpus processing operations (merge, split, search) have been written and a LaTeX/PGF graphical output mechanism has been added in order to expertise and analyse the content of annotated utterances. This visualization functionality has been developed during the 3-month summer internship of Andrei Zene, a Romanian B.Sc. student.

The toolkit Roots is original compared to other related tools. Among them, GATE [CB02] proposes a framework to develop NLP pipelines but does not provide facilities to switch between GATE bundled processing components and external tools. More recently, the NITE XML Toolkit, or NXT, proposes a generic data organization model able to represent large multimodal corpora with a wide range of annotation types [CEHK05,CCB+10]. Whereas NXT considers corpora as databases from which data is accessed through a query language, Roots lets the user browse data as he sees fit. In a more general approach, UIMA [FL04,FLG+06] proposes software engineering standards for unstructured data management, including annotation and processing. UIMA is technically too advanced for fast and light prototyping. It is rather devoted to industrial developments. In the end, Roots is closer to work done within the TTS system Festival [BTCC02]. This system relies on a formalism called HRG, standing for Heterogenous Relation Graphs, which offers a unique representation of different information levels involved in the TTS system [TBC01]. Still, our tool is different from HRG in the sense that the latter is part of the TTS system Festival whereas Roots is completely autonomous. Moreover, Roots comes along with a true application programming interface (API), in C++ and Perl for the moment.

As a result of recent improvements, Roots is now in use in most of the software developed for speech processing, namely the corpus-based speech synthesizer, corpus generation/analysis tools or the phonetizer. Moreover, Roots serves as a basis for corpus generation and information extraction for the ANR Phorevox project. For instance, we have built a corpus containing 1000 free e-books which is planned to be proposed to the community. Finally, Roots has been registered in 2013 at the Program Protection Agency (*Agence pour la Protection des Programmes*, APP) and publicly released under the terms of LGLP licence on `http://roots-toolkit.gforge.inria.fr`. A paper has been published in the main international language resource conference to let the community know about this release [CLL14].

[CB02]   D. Cunningham, H.and Maynard, V. Bontcheva, K.and Tablan, "GATE: an architecture for development of robust HLT applications", *in : Proceedings of the Annual Meeting of the ACL*, p. 168–175, 2002.

[CEHK05]   J. Carletta, S. Evert, U. Heid, J. Kilgour, "The NITE XML Toolkit: Data Model and Query Language", *Language Resources and Evaluation 39*, 4, 2005, p. 313–334.

[CCB+10]   S. Calhoun, J. Carletta, J. M. Brenier, N. Mayo, D. Jurafsky, M. Steedman, D. Beaver, "The NXT-format Switchboard Corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue", *Language Resources and Evaluation 44*, 4, 2010, p. 387–419.

[FL04]   D. Ferrucci, A. Lally, "UIMA: an architectural approach to unstructured information processing in the corporate research environment", *Natural Language Engineering 10*, 3-4, 2004, p. 327–348.

[FLG+06]   D. Ferrucci, A. Lally, D. Gruhl, E. Epstein, M. Schor, J. W. Murdock, A. Frenkiel, E. W. Brown, T. Hampp, Y. Doganata *et al.*, "Towards an interoperability standard for text and multi-modal analytics", *research report*, 2006.

[BTCC02]   A. W. Black, P. Taylor, R. Caley, R. Clark, "The Festival speech synthesis system", *research report*, University of Edinburgh, 2002.

[TBC01]   P. Taylor, A. W. Black, R. Caley, "Heterogeneous relation graphs as a formalism for representing linguistic information", *Speech communication 33*, 2001, p. 153–174.

[CLL14]   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, May 2014, `https://hal.inria.fr/hal-00974628`.

4.3   Web-based listening test system

Participants: Vincent Barreaud, Arnaud Delhay, Sébastien Le Maguer, Damien Lolive.

The listening test platform is developed by the team, especially to evaluate speech synthesis models. This platform has been developed to propose to the community a ready to use tool to conduct listening tests under various conditions. Our main goals were to make the configuration of the tests as simple and flexible as possible, to simplify the recruiting of the testees and, of course, to keep track of the results using a relational database.

The most widely used listening tests used in the speech processing community are available (AB-BA, ABX, MOS, MUSHRA, etc.).

This software is currently implemented in PHP and integrated in the Symfony2 framework with Doctrine as database manager and Twig templates. This configuration makes the platform accessible from a wide variety of browsers.

The platform is designed to enable researchers to build wide tests available through the web. The main functionalities provided are as follows:

- Users are given roles, which give them privileges,

- Users answer test during a trial which can be interrupted and resumed later,

- Users give information on their listening conditions at each trial beginning,

- Tests are imported from Zip archives that contain a XML configuration file and the stimulii,

- Users can be imported from a XML configuration file.

- A tester can monitor his test and discard results of a testee on the basis of its statistical behavior.

- The platform is open-source (under AGPLv3 Licence).

4.4   Automatic segmentation system

Participants: Damien Lolive.

The automatic segmentation system consists of a set of scripts aligning the phonetic transcription of a text with its acoustic signal counterpart. The system is made of two parts: the first one generates a phonetic graph including phonological variants (pauses, liaisons, schwas,...), the second one, based on HMM modeling, performs a Viterbi decoding determining the most likely phonetic sequence and its alignment with the acoustic signal.

To be efficient, the system should be applied to texts that have been manually checked (compliance with the recording, spelling, syntax) and annotated. The annotation stage consists in adding tags indicating excepts in foreign language, non standard pronunciation and noises (breathing, laughter, coughing, sniffing, snorting, etc.). It is also possible to improve the decoding performances by adding a list of phonetization of proper names and foreign pronunciations.

4.5   Corpus-based Text-to-Speech System

Participants: Nelly Barbot, Jonathan Chevelu, Arnaud Delhay, David Guennec, Damien Lolive.

For research purposes we developed a whole text-to-speech system designed to be flexible. The system, implemented in C++, intensively use templates and inheritance, thus providing the following benefits:

- the algorithm used for unit selection can be easily changed. For instance, we implemented both $A^\star$ and Beam-search simply by using subclassing and without changing the heart of the system.

- cost functions can also be changed the same way which provides a simple way to experiment new functions.

Moreover the system implements state of the art technique to achieve good performance while manipulated large speech corpora such as hash tables and pre-selection filters. To achieve this, each phone in the corpus is given a binary key which enables A^* to take or reject the unit. Thus, the key contains phonetic, linguistic and prosodic information. Binary masks are used to get access only to the desired information during runtime.

The pre-selection filters are integrated to the hash functions used to access the units in the corpus in order to reduce the number of candidates explored. For the moment, the whole set of filters is the following:

1. Is the unit a Non Speech Sound ?

2. Is the phone in the onset of the syllable?

3. Is the phone in the coda of the syllable?

4. Is the phone in the last syllable of its breath group?

5. Is the current syllable in word end?

6. Is the current syllable in word beginning?

Concretely, the pre-selection filters are relaxed one by one, starting from the end of the list, if no unit corresponding to the current set is found. One drawback is that we can explore candidates far from the target features we want, thus risking to produce artefacts but this backtracking mechanism insures to find a unit and to produce a solution. The priority order of the filters is the one given above.

Finally, high level features are also available to get, for example, the best path or the N-best paths, with a detailed output of the cost values.

Some developments are currently undertaken to provide more features and pre-selection filters and also to improve flexibility of the system to gain a fine control over prosody. This last objective is linked to the main objectives of the team to control expressivity during synthesis.

4.6   Recording Studio

Participants: Nelly Barbot, Vincent Barreaud, Damien Lolive.

A main goal of the EXPRESSION project consists in developing high quality voice synthesis. Our research activities use speech corpora as a raw material to train statistical models. A good speech corpus quality relies on a consistent speech flow (the actor does not change his speaking style during a session) recorded in a consistent (and quiet) acoustic environment. In order to expand our research scope, it is often interesting to vary the speech style (dialogs, mood, accent, etc.) as well as the language style. Unfortunately, such corpora are hard to obtain and generally do not meet specific experimental requirements. To deal with these constraints, speech resources need to be recorded and controlled by our own protocols.

4.6.1 Hardware architecture

The funding of this recording studio comes from MOB-ITS (CPER, 2007-2013). The MOB-ITS platform (Mobile and interactive access to data) is a joint project of IRISA teams in Lannion (IUT and ENSSAT). This contract is part of the support to the "Pôle de compétitivité Images & Réseaux".

This recording studio consists in two rooms: an isolation booth and control room.

The isolation booth can fit three persons. It is designed to attenuate the noises of 50dB and is equipped with two recording sets. A recording set consists in a high quality microphone (Neumann U87AI), a high quality closed head set (Beyer DT 880 250ohms), a monitor and a webcam.

The control room is equipped with two audio networks, a video network and computer network. The first audio network is a high quality digital recording line going from the isolation booth microphones to a digital sound card through a preamplifier (Avalon Design AD2022), an equalizer (Neve 8803 Dual Channel) and finally an analogic/digital converter (Lynx Aurora 8). The digital sound is edited with a logical sampling table (Avid Pro Tools).

In addition to the signal issued by the isolation room, the digital audio network can record the signals from an Electro-Gloto Graph (EGG) that capture the glottal activity of the actor. This activity is used to induce the F0 (first formant) trajectory which is the main indicator of the prosody. This activity must be digitalized and recorded along with the audio activity in order to reduce the latency between the two signal.

The second audio network is for control purpose and is fully analogic. It is used by the operator to control the quality of the recorded sound, the consistency of the actor, the accuracy of the transcription. An actor can receive audio feedback of his own voice, disturbing stimulii (music, other voices, their own delayed voice) or directions from the operator through this audio line. This network consists in four Neumann KH 120 loud-speakers (two in the booth, two in the control room), a head set amplifier (ART headamp 6 pro) and an analogic sampling table (Yamaha MG206C). The computer network stores the recording sessions scenarii and prompt the actor.

The video network switches the video output (computers, webcam) to screens installed in the isolation booth (for prompting) and the control room (for monitoring).

4.6.2 Software architecture

Actors in the isolation booth must be prompted to utter speech with various indications (mood, intonation, speed, accent, role, ...). The prompt must be presented on the simplest interface, for instance a lcd screen or a tablet. The latest developments on the recording studio consist in a software implemented on the computer at the end of the digital audio network that record sound files, segment them and link them to the transcription. This software is controlled by the operator who checks that the actors actually uttered the prompted sentence and the quality of the recording. Thus, the operator can possibly reject (in fact, annotate) a file and prompt the actors again with the discarded sentence.

The digital sound card used for recording only offers microsoft drivers. Consequently, this software has been developed with the Windows Audio and Sound API (WASAPI). The main difficulty resides in the simultaneous recording of two distinct channels (audio and EGG) without any jitter between the two signals.

5 Contracts and Grants with Industry

5.1 INGREDIBLE

Participants: Pamela Carreño, Sylvie Gibet, Pierre-François Marteau.

The *INGREDIBLE* project project is funded by the French Research Agency (program ANR

CONTINT. The partners are: LabSTICC (team leader), LIMSI-CNRS, IRISA, Virtualys, Final users (DEREZO, Brest; STAPS lab., Orsay.

The goal of the INGREDIBLE project is to propose a set of scientific innovations in the domain of human/virtual agent interaction. The project aims to model and animate an autonomous virtual character whose bodily affective behavior is linked to the behavior of a human actor. The outcomes of the project are:

- The creation of different motion corpora (fitness, interactive video games motions, theatrical gestures);

- The development of algorithms for affect recognition from expressive features;

- The development of a behavioral system linked to the recognition and synthesis modules;

- The development of synthesis algorithms for animating a virtual character in interactive situations.

5.2 SynPaFlex

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

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.

6   Other Grants and Activities

6.1   International Collaborations

- David Guennec has visited the Idiap Research Institute in Martigny, Switzerland during 5 months in 2015. During his stay, he worked on prosody modelling and integration into the speech synthesis process. For instance, this collaboration with Philip N. Garner at Idiap enabled to introduce the atom-based model, which relies on the Fujisaki model, into unit selection. A paper has been submitted to the IEEE Signal Processing Letters.

6.2   National Collaborations

- **Hybride ANR Project** Participant: Nicolas Béchet. The Hybride Research Project aims at developing new methods and tools for supporting knowledge discovery from textual data by combining methods from Natural Language Processing (NLP) and Knowledge Discovery in Databases (KDD). The consortium is made of INRIA/LORIA, GREYC, MoDyCo, Inserm (IRISA is associated to the GREYC in this project). The coordinator is Yannick Toussaint from INRIA/LORIA.

- **ANIMITEX CNRS MASTODONS Project** Participant: Nicolas Béchet. The ANIMITEX Project aims is to exploit the massive and heterogeneous textual data to provide crucial information in order to complete the analysis of satellite images. The consortium is made of LIRMM, TETIS, ICUBE, GREYC, LIUPPA, IRISA. The coordinator is Mathieu Roche, Cirad/TETIS.

7 Dissemination

7.1 Involvement in the Scientific Community

- Pierre-François Marteau served as a reviewer in international journals (IEEE TPAMI, IEEE TNNLS, IEEE TKDE, PRL). He serves as an expert for French Ministry of Research (CIR/JEI) and ANRT (CIFRE). He was member of a thesis committee at Nantes University, LINA. He is member of the Strategic Orientation Committee at IRISA and member of the scientific committee at Université de Bretagne Sud.

- Sylvie Gibet serves as a reviewer for the IEEE Transactions on Affective Computing, the IEEE Transactions on Pattern Analysis and Machine Intelligence. She has served as a reviewer for the second Workshop on Motion Computing (MOCO). She has been president of the jury for the thesis of Billal Merabti (IRISA, Rennes, 2015).

- Arnaud Delhay has been re-elected as a member of the 'Commission Recherche' (Research comittee) of the IUT of Lannion in November 2015.

- Damien Lolive is an elected member of the 'Conseil Scientifique' (Scientific council) of ENSSAT, Lannion. He also 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) and for the "*Journées d'Études sur la Parole*".

- Gwénolé Lecorvé is an elected member of the laboratory council of IRISA. He also serves as a reviewer for the International conference of the International Speech Communication Association (Interspeech), and for the "*Journées d'Études sur la Parole*".

- Caroline Larboulette is a member of various program committees for international conferences (CAe/EXPRESSIVE 2015, CASA 2015), 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). She is a member of the ACM SIGGRAPH Small Conferences Committee that attributes the ACM SIGGRAPH labels to conferences and supervises the budget of conferences sponsored by ACM SIGGRAPH. She is also vice-president of the ACM SIGGRAPH Madrid Professional Chapter that she created in 2007.

7.2 Teaching

- 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).

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

- Sylvie Gibet teaches the following computer science courses at the faculty of sciences, Université de Bretagne Sud: functional programming and algorithmic in License level, and multimedia digital processing and motion computing in Master level (master WMR).

- 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.

- 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 logic for undergraduates of the UFR SSI; C++ programming for ENSIBS graduate students; and computer animation, rendering and interactive techniques at the master level (Master WMR).

- Gwénolé Lecorvé teaches the following computer science courses at École Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT): Java, graphical interfaces, C++; distributed algorithmics; and artificial intelligence in Master level.

- 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, and XML, compilers architecture and formal languages theory in Master Level.

- Pierre-François Marteau teaches programming languages, logics, introduction to cryptography and information retrieval 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

- Jeanne Villaneau teaches various computer sciences courses at École Nationale Supérieure de Bretagne Sud.

7.3   Conferences, workshops and meetings, invitations

- Pierre-François Marteau has co-organized with Ahlame Douzal (LIG, Grenoble Alpes University, France), and José A. Vilar (University A Coruna, Spain) the first ECML/PKDD 2015 Workshop on: Advanced Analytics and Learning on Temporal Data (AALTD'2015) Friday, September 11, 2015, Porto, Portugal. 20 papers have been presented orally or as posters, and a post-act volume in Springer LNAI series is in preparation.

7.4   Graduate Student and Student intern

- Cédric Fayet has done is Master level internship in Lannion. He has worked on the language independency of the TTS system using machine learning techniques. The main goal of his work was to predict phoneme segments based on phonologic and positional features instead of using phonetizer and language dependent labels.

8   Bibliography

Major publications by the team in recent years

[1] N. Barbot, V. Barreaud, O. Boëffard, L. Charonnat, A. Delhay, S. Le Maguer, D. Lolive, "Towards a Versatile Multi-Layered Description of Speech Corpora Using Algebraic Relations", *in : Proceedings of Interspeech*, p. 1501–1504, 2011.

[2] 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.

[3] 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`.

[4] G. Ke, P.-F. Marteau, G. Ménier, "Improving the clustering or categorization of bi-lingual data by means of comparability mapping", October 2013.

[5] 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.

[6] P.-F. Marteau, S. Gibet, "On Recursive Edit Distance Kernels with Application to Time Series Classification", February 2013.

Articles in referred journals and book chapters

[7] N. Barbot, O. Boëffard, J. Chevelu, A. Delhay, "Large Linguistic Corpus Reduction with SCP algorithms", *Computational Linguistics 41*, 3, September 2015, p. 30, `https://hal.inria.fr/hal-01135089`.

[8] S. Gibet, F. Lefebvre-Albaret, L. Hamon, R. Brun, A. Turki, "Interactive editing in French Sign Language dedicated to virtual signers: requirements and challenges", `https://hal.archives-ouvertes.fr/hal-01205742`.

[9] H. Yoo, E. Delais-Roussarie, D. Lolive, N. Barbot, "Le Rythme en Lecture Oralisée (parole synthétique et parole naturelle)", *Revue française de linguistique appliquée XX*, 2, 2015, p. 63–77, `https://hal.inria.fr/hal-01272551`.

Publications in Conferences and Workshops

[10] P. Alain, J. Chevelu, D. Guennec, G. Lecorvé, D. Lolive, "The IRISA Text-To-Speech System for the Blizzard Challenge 2015", *in : Blizzard Challenge 2015 Workshop*, p. 4 p., 2 columns, Berlin, Germany, September 2015, `https://hal.inria.fr/hal-01196168`.

[11] 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`.

[12] J. Chevelu, D. Lolive, "Do not build your TTS training corpus randomly", *in : Proceedings of the European Signal Processing Conference (EUSIPCO)*, Nice, France, August 2015, `https://hal.inria.fr/hal-01199083`.

[13] M. Dupont, P.-F. Marteau, "Coarse-DTW: Exploiting Sparsity in Gesture Time Series", *in : Advanced Analytics and Learning on Temporal Data (AALTD)*, A. Douzal-Chouakria *et al.* (editors), *Proceedings of the 1st International Workshop on Advanced Analytics and Learning on Temporal Data (AALTD)*, *CEUR Vol-1425*, CEUR Workshop Proceedings, Porto, Portugal, September 2015, `https://hal.inria.fr/hal-01205740`.

[14] M. Dupont, P.-F. Marteau, "Gesture control system for mobile robots", *in : 27ème conférence francophone sur l'Interaction Homme-Machine.*, ACM, p. w20, Toulouse, France, October 2015, `https://hal.archives-ouvertes.fr/hal-01219123`.

[15] D. Guennec, J. Chevelu, D. Lolive, "Defining a Global Adaptive Duration Target Cost for Unit Selection Speech Synthesis", *in : International Conference on Text, Speech and Dialogue (TSD)*, *Proceedings of International Conference on Text, Speech and Dialogue (TSD)*, p. 157–165, PLZEŇ, Czech Republic, September 2015, `https://hal.inria.fr/hal-01188686`.

[16] M. Harzallah, G. Berio, "A unified framework for semantic comparison of objects: extension to semantic graph comparison", *in : 19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES2015)*, Singapour, Singapore, September 2015, `https://hal.archives-ouvertes.fr/hal-01165148`.

[17] C. Larboulette, S. Gibet, "A Review of Computable Expressive Descriptors of Human Motion", *in : MOCO 2015 : 2nd International Workshop on Movement and Computing*, *Proceedings of the International Workshop on Movement and Computing*, p. 21–28, Vancouver, Canada, August 2015, `https://hal.archives-ouvertes.fr/hal-01196267`.

[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] P.-F. Marteau, "Noyaux temporellement élastiques régularisés et ACP non-linéaire", *in : EGC 2015*, T. T. Jérome Darmont, Benoît Otjacques (editor), *EGC 2015. Revue des Nouvelles Technologies de l'Information*, *RNTI-E-28*, RNTI, p. 47–58, Luxembourg, Luxembourg, January 2015, `https://hal.archives-ouvertes.fr/hal-01111658`.

[20] 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`.

[21] C. Reverdy, S. Gibet, C. Larboulette, "Animation faciale basée données : un état de l'art", *in : 28èmes journées de l'Association Française en Informatique Graphique*, *Proceedings of AFIG*, Lyon, France, November 2015, `https://hal.archives-ouvertes.fr/hal-01366447`.

[22] C. Reverdy, S. Gibet, C. Larboulette, "Optimal Marker Set for Motion Capture of Dynamical Facial Expressions", *in : MIG2015 8th ACM SIGGRAPH Conference on Motion in Games*, *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*, p. 31–36, Paris, France, November 2015, `https://hal.archives-ouvertes.fr/hal-01279483`.

[23] H.-H. Vu, J. Villaneau, F. Saïd, P.-F. Marteau, "Mesurer la similarité entre phrases grâce à Wikipédia en utilisant une indexation aléatoire", *in : TALN 2015*, Caen, France, June 2015, `https://hal.archives-ouvertes.fr/hal-01167929`.

Miscellaneous

[24] M. Harzallah, G. Berio, "Towards an approach for configuring ontology validation", working paper or preprint, February 2015, `https://hal.archives-ouvertes.fr/hal-01118953`.

[25] P.-F. Marteau, "Times series averaging from a probabilistic interpretation of time-elastic kernel", working paper or preprint, May 2015, `https://hal.archives-ouvertes.fr/hal-01155134`.