EXPRESSION Team
Expressiveness in Human Centered Data/Media

Long version - V2

22/10/2014

Contents

1 Team composition . . . 3
2 Introduction . . . 4
3 Main challenges and research focus . . . 4
3.1 Main challenges addressed by the team . . . 4
3.2 Main research focus . . . 6
4 Expressive gesture analysis, synthesis and recognition . . . 7
4.1 Scientific background . . . 7
4.2 Challenges . . . 8
4.3 Research topics . . . 9
4.4 Application areas . . . 10
4.5 Key expected results . . . 11
5 Expressive speech analysis and synthesis . . . 11
5.1 Scientific background . . . 11
5.2 Challenges . . . 13
5.3 Research topics . . . 13
5.4 Application areas . . . 15
5.5 Key expected results . . . 16
6 Expressiveness in textual data . . . 16
6.1 Scientific background and challenges . . . 17
6.2 Research topics . . . 19
6.3 Application areas . . . 21
6.4 Key expected results . . . 21
7 Main expected outcomes and synergies between modalities . . . 21
8 Positioning . . . 22
8.1 National and international related teams and laboratories . . . 23
8.2 Related teams at IRISA . . . 26
8.3 Strategy to develop visibility and impact . . . 27
9 Team experience . . . 27
9.1 National on-going projects . . . 27
9.2 International on-going projects . . . 29
9.3 Industrial cooperation . . . 30
9.4 Defended PhDs . . . 30
9.5 Software . . . 30
9.6 Technical facilities and platforms . . . 32
9.7 Corpora . . . 34
10 Brief risk analysis . . . 34
11 Appendix: team environment . . . 35
11.1 National environment . . . 35

11.2 International environment 37

1 Team composition

Team leader			
Pierre-François Marteau	PR*	UBS†	computer science, text, gesture
Permanent staff			
Nelly Barbot	MCF*	UR1†	applied mathematics, speech
Nicolas Béchet	MCF	UBS	computer science, text
Giuseppe Bério	PR	UBS	computer science, text
Arnaud Delhay	MCF	UR1	computer science, speech
Sylvie Gibet	PR	UBS	computer science, gesture
Jean-François Kamp	MCF	UBS	computer science, gesture
Gwénolé Lecorvé	MCF	UR1	computer science, speech, text
Damien Lolive	MCF	UR1	computer science, speech
Gildas Ménier	MCF	UBS	computer science, text, gesture
Jeanne Villaneau	MCF	UBS	computer science, text
Associated staff			
Vincent Barreaud	MCF	UR1	computer science, speech
Farida Said	MCF	UBS	applied mathematics, text
Caroline Larboulette	PhD	Animation, computer science	
PhD candidates			
Pamela Carreno	3rd year	Expressive gesture synthesis, ANR INGREDIBLE	
Lei Chen	3rd year	Gesture interaction, with univ. McGill, Montréal, 1/2 Brittany regional grant	
Marc Dupont	1st year	Gesture recognition in mobility, with THALES TOSA	
David Guennec	3rd year	Corpus-based speech synthesis	
Hai Hieu Vu	3rd year	Text mining, Vietnamese grant	
Raheel Qader	1st year	Pronunciation modeling for speech synthesis	
Clément Reverdy	1st year	Synthesis and recognition of facial expressions	
Temporary staff			
Jonathan Chevelu	IGR	09/2012	ANR Phorevox
Ludovic Hamon	post-doc	09/2012	SIGN3D project, computer science
Recently defended PhDs			
Larbi Mesbahi	12/2010	Voice conversion	
Charly Awad	02/2011	Movement synthesis for animation	
M. Marwan M. Fuad	02/2011	Time series mining	
Ismaïl El Maarouf	12/2011	Semantic relation extraction from textual data	
Kyle Duarte	06/2012	Sign language analysis and synthesis	
Sébastien Le Maguer	07/2013	Statistical speech synthesis	
Thibault Le Naour	12/2013	Virtual characters animation	
Guyao Ke	02/2014	Comparable corpora and text mining	

*PR: full professor; MCF: associate professor.

†UBS: Université de Bretagne Sud; UR1: Université de Rennes 1.

2 Introduction

Expressiveness or expressivity are terms which are often used in a number of domains[1]. When it comes to human expressiveness, we will consider the following reading: expressiveness covers any kind of variability -produced by a human being during the act of producing a semantically rich *content*- in particular variability that conveys emotion, style or intentional content.

Considering this definition, the EXPRESSION team focuses on studying human language data conveyed by different media: gesture, speech and text as depicted in figure 3.1. 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 numerical (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 and expressiveness, 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. Expressiveness in textual data.

3 Main challenges and research focus

3.1 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[2] (metric or similarity space embeddings) in order to ease post-processing of data,

[1] 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.

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

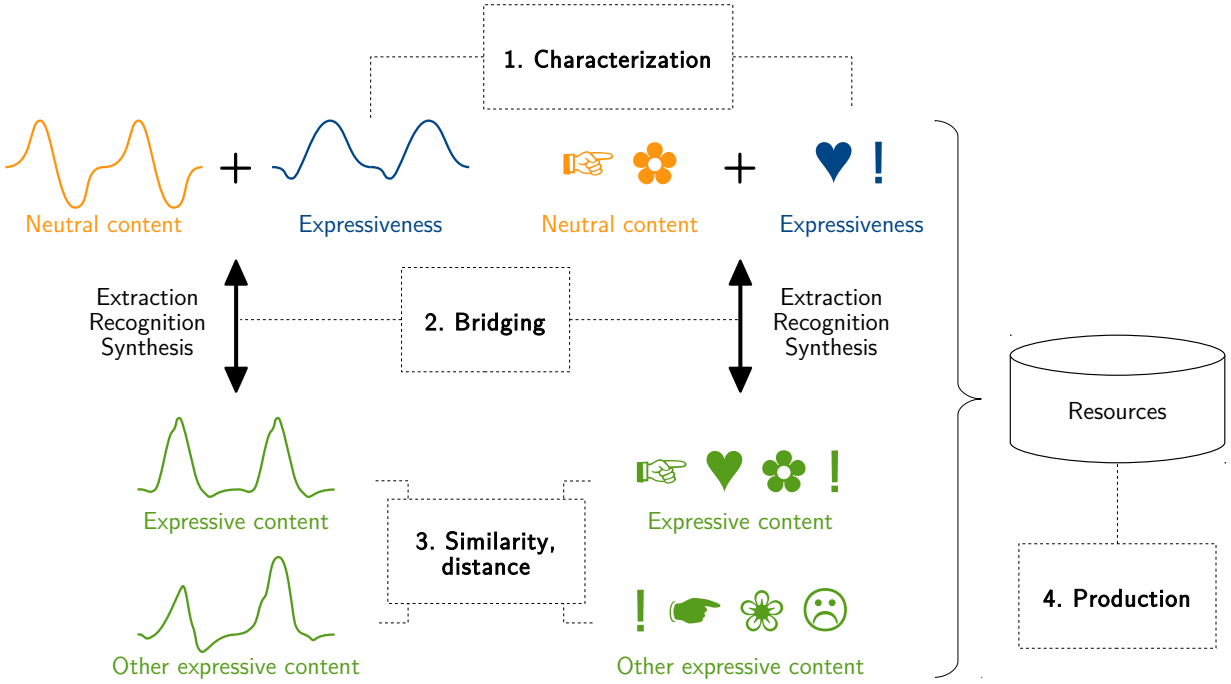

Figure 1: Overview of the main challenges considering both on continuous numerical (left) and discrete symbolic (right) data.

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.

3.2 Main research focus

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

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

RF2: Multi-level representations – We rely on multi-level representations (semantic, phonological, phonetic, signal processing) to organize and apprehend data. The heterogeneity of these representations (from metadata to raw data) prevents us from using standard modeling techniques that rely on homogeneous features. Building new multi-level representations is thus a main research direction. Such representations will provide efficient information access, support for database enrichment through bootstrapping and automatic annotation. This research focus contributes mainly to the second, third and fourth challenges;

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

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

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

of cross-disciplinary evaluation methodologies. This research focus contributes also to the second challenge;

4 Expressive gesture analysis, synthesis and recognition

4.1 Scientific background

Thanks to advanced technologies such as new sensors, mobile devices, or specialized interactive systems, gesture communication and expression have brought a new dimension to a broad range of applications never before experienced, such as entertainments, pedagogical and artistic applications, rehabilitation, etc. The study of gestures requires more and more understanding of the different levels of representation underlying their production, from meanings to motion performances characterized by high-dimensional time-series data. This is even more true for skilled and expressive gestures, or for communicative gestures, involving high level semiotic and cognitive representations, and requiring extreme rapidity, accuracy, and physical engagement with the environment.

Many previous works have studied movements and gestures that convey a specific meaning, also called semiotic gestures. In the domain of co-verbal gestures, Kendon [Ken80] is the first author to propose a typology of semiotic acts. McNeil extends this typology with a theory gathering the two forms of expression, speech and action [McN92]. In these studies, both modalities are closely linked, since they share a common cognitive representation. Our research objectives focus more specifically on body movements and their different forms of variations in nonverbal communication or bodily expression. We consider more specifically full-body voluntary movements which draw the user's attention, and express through body language some meaningful intent, such as sign language or theatrical gestures. Generally, these movements are composed of multimodal actions that reveal a certain expressiveness, whether unintentional or deliberate.

Different qualitative aspects of expressiveness have already been highlighted in motion. Some of them rely on the observation of human motion, such as those based on the Laban Movement Analysis theory, in which the expressiveness is essentially contained into the Effort and Shape components [Mal87]. Motion perception through bodily expressions has also given rise to many work in nonverbal communication. In the psychology and neuroscience literature, recent studies have focused in particular on the recognition of emotion in whole body movements [Wal98; Gal09; THB06; dGel06; CG07].

In computational sciences, many studies have been conducted to synthesize expressive or emotional states through the nonverbal behavior of expressive virtual characters. Two major classes of approaches can be distinguished: those that specify explicit behaviors associated with pure synthesis techniques, or those offering data-driven animation techniques. In the first category we find embodied conversational agents (ECAs) that rely on behavioral description languages [KW04], or on sets of expressive control parameters [CCZB00; HMBP05]. More recently, some computational models consider the coordination and adaptation of the virtual agent with a human or with the environment in interacting situations. The models in such cases focus on rule-based approaches derived from social communicative theories [Pel09; Kop10]. In the second category, motion captured data is used with machine learning techniques to capture style in motion and generate new motion with variations in style [BH00; Her03; GMHP04; HPP05]. In these works authors consider a low-level definition of style, in terms of variability observed among several realizations of the

same gesture. If some relevant studies rely on qualitative or quantitative annotations of motion clips (*e.g.*, [AFO03; MBS09]), or propose relevant methods to create a repertoire of expressive behaviors (*e.g.*, [RBC98]), very few approaches deal with both motion-captured data and their implicit semantic and expressive content.

In our approach, we will consider that gesture is defined as expressive, meaningful bodily motion. It combines multiple elements which intrinsically associate *meaning*, *style*, and *expressiveness*. The *meaning* is characterized by a set of signs that can be linguistic elements or significant actions. This is the case when gestures are produced in the context of narrative scenarios, or expressive utterances in sign languages. The *style* includes both the identity of the subject, determined by the morphology of the skeleton, the gender, the personality, and the way the motion is performed, according to some specific task (e.g., moving in a graceful or jerky way). The *expressiveness* characterizes the nuances that are superimposed on motion, guided by the emotional state of the actor, or associated to some willful intent. For example, theatrical performances may contain intentional emphasis that are accompanied by effects on the movement kinematics or dynamics. Most of the time, it is very difficult to separate all these components, and the resulting movements give rise to different physical realizations characterized by some variability that can be observed into the raw motion data and subsequently characterized. For simplicity we will assume later that the notion of expressiveness includes any kind of variability.

Hence our line of research focuses specifically on the study of variability and variation in motion captured data, linked to different forms of expressiveness, or to the sequencing of semantic actions according to selected scenarios. Motion capture is used for retrieving relevant features that encode the main spatio-temporal characteristics of gestures: low-level features are extracted from the raw data, whereas high-level features reflect structural patterns encoding linguistic aspects of gestures [ACD+09]. Many data-driven synthesis model have been developed in order to re-use or modify motion capture data and therefore produce new motions with all the realism and nuances present in the examples. We focus in our approach on machine learning methods that capture all the subtleties of human movement and generate more expert gestures while maintaining the style, expressiveness and semantic inherent to human actions [Her03; AI06; HCGM06b; PP10]. One of the novelties of our approach is that it is conducted through an analysis / synthesis scheme, corrected and refined through an evaluation loop (*e.g.*, [GMD12]). Consequently, data-driven models, which incorporate constraints derived from observations, should significantly improve the quality and credibility of the gesture synthesis; furthermore, the analysis of the original or synthesized data by techniques of automatic segmentation, classification, or recognition models should improve the generation process, for example by refining the annotation and cutting movements into significant items. Finally, evaluation takes place at different levels in the analysis / synthesis loop, and is performed qualitatively or quantitatively through the definition of original use cases.

4.2 Challenges

The first challenge is to define and characterize the expressiveness in human movement. As stated above, we will consider expressiveness at all levels of gesture generation, which involves both a semantic dimension (from actions that convey a specific meaning to sign languages that imply the linguistic aspects of phonetics, phonology, prosody, etc.), and an expressive dimension induced by intentional variations or spontaneous states of the actor, and results in variations in the signals [HCGM06a]. This challenge is part of the general challenge C1.

The second challenge is to explore new motion representation spaces that reflect the expressiveness and variability contained in the data. This implies to reduce the complexity of the high-dimensional motion data by proposing different embeddings for these data [WFH07; CL06; EL13; MHM12; PT09]. Such embeddings should enable to characterize and parametrize specific action sequences, and give rise to original approaches for recognition or generation of new behaviors with varied styles, applied to expressive movements and gestures. This challenge relates to the challenge C3.

The third challenge is to be able to link the different levels of representation, from narrative scenarios to structural patterns of actions, and to continuous streams of raw motion data. More precisely, the aim is to extract structural patterns from data and to understand how these discrete patterns influence the synthesis of gestures while preserving the semantics of actions as well as subtle expressive variations [RBC98; GCDL11]. This challenge deals with the general challenge C2.

The fourth challenge concerns the definition of evaluation protocols that are necessary for evaluating the different hypothesis and models that are constructed at all the levels of the perception-production loop. Our approach follows the motor theory of perception, where motor production is necessarily involved in the recognition of sensory cues (audio, visual, etc.) and encoded actions [GMD12]. These evaluations will be considered both quantitatively and perceptually, following original methodologies that take into account the scenarios, movement tasks, elicitation and recognition [GCF10; BR09]. This challenge contributes to the general challenges C1 and C4.

4.3 Research topics

Scenarios and corpora: As the expressiveness in gestures is a notion still not well established and somehow controversial, its characterization implies the definition of relevant scenarios incorporating some variability induced by the specification of tasks and the pre-determination of different sources of variations (by varying scenarios, setting up variational emotional states, etc.). For example, an important requirement for most signing avatar technology is to ensure that the constructed database has a large quantity of interesting and on-topic signs from which to build novel signing sequences. Key factors such as dialogue content and style, as well as linguistic inclusions, have to be considered in designing an avatar corpus [DG10].

When these scenarios and corpora have been designed, motion databases are constructed. This research topic is part of the RF1 research focus. The following thematic research axis that rely on these databases are identified to carry out our research on gesture.

Multi-level representations: expressive gestures rely on multi-level representations, depending on the type of gesture and task that are considered. For sign languages, we may identify levels (semantic, phonological, phonetic, or signal processing) to organize and process the data. In addition, the data is organized spatially and temporally. The spatial dimension is characterized by channels associated to some body parts (hands, arm movements, facial expression, gaze direction, etc.), whereas the time dimension is characterized by segmental elements that may represent significant phonetic items, signs or actions, or motion transitions. Associated to this information are expressive features that bring additional dimensions to the description of the motion data. For example, for narrative scenes in theater, different action sequences can be performed with different emotional states, or for musical scores, different gestural interpretations can be performed with various nuances or musicality. Defining the features that best characterize these expressive sequences is still challenging, as highlighted by [AC14] who automatically compute motion qualities from dance performances, in terms of Laban Movement Analysis (LMA), for motion analysis and indexing purposes. Finding the most adequate multi-level representations and the way to

index motion data according to these heterogenous information, both discrete (semantic labeling) and continuous (e.g., kinematic or spatial features) is thus a second research direction. Such representations will provide efficient information access, support for database enrichment through boostrapping and data generation, and automatic segmentation and annotation. This research topic relates to the RF2 research focus.

Knowledge extraction, Analysis, Classification: based on the previous data representation, this research addresses motion signal processing (filtering, retrieving, clustering, classification, recognition). As motion is by essence constituted of high-dimensional time series, the data-driven models that are developed will be mostly based on machine learning methods and will consider the definition of relevant low-dimensional embeddings in which the data will be projected [CGM14]. The characterization of distances or similarity measures in such embeddings is also a key issue. One idea is to define efficient embeddings that may have a sensori-motor plausibility, such that the resulting parametrization will be significant and will make possible a better generalization of the data. This research topic is part of the RF3 research focus.

Generation and control: our fourth research axis is related to the automatic generation of high-quality movements gestures mimicking or extrapolating new human behaviors. In addition, to guarantee adequate expressiveness, the variability of the produced gestures has to be finely controlled. The methods are based on data-driven approaches and on cognitive and machine learning control processes. For some autonomous applications, such as the production of utterances by signing avatars, the statements and actions can be generated from structural models (composition of gestures in French sign language (LSF) from parametrized linguistic units). Other interactive 3D applications are directly controlled by gestures that are tracked or recognized in real-time. This research topic is included in the RF4 research focus.

Use cases and evaluation: controlling dynamical systems driven by gesture and producing sensory outputs (gestural, visual, auditive) requires final perceptual (i.e., manual) evaluations. These evaluations focus mainly on the expressive quality of the gestures that are recognized and produced, and therefore they suppose to adapt classical evaluation methodologies to novel approaches. Furthermore, they are closely linked to the use cases previously stated. Hence, the evaluation results can be used to refine the different analysis and synthesis models, or to recreate new use cases through a closed loop approach which is relevant to characterize more finely the notion of expressiveness in motion. This research topic contributes to the RF5 research focus.

4.4 Application areas

Different applications domains can be considered:

- Sign Language Translation and Avatar technology; This application domain covers in particular the design of corpora and sign language indexed databases, the development of analysis / synthesis software to control sign language virtual characters [GCDL11], and the design of innovative interfaces to manipulate the data. This kind of application may require the recording of high-quality data (body and hand motion, facial expression, gaze direction), or real-time interactive devices to communicate more efficiently and intuitively with the application. Sign language video books can be a targeted application.
- Interactive Multimedia Technology using Gesture; Controlling expressively by gesture the behavior of simulated objects is an emerging research field which can lead to numerous

applications: games using gesture as input or virtual assistants as output, virtual theater, or more generally performative art controlled by gesture.

4.5 Key expected results

- Models of expressiveness in gesture both for recognition and generation.
- High-quality expressive gesture generation from motion captured data, using both low-level embeddings that implicitly integrate expressive descriptors, and high level semantic patterns.
- Development of new similarity measures for motion data, efficient classification, recognition and generation algorithms based on kernel-based techniques.
- Construction of novel scenarios and corpora that enable to explore various expressive situations.

5 Expressive speech analysis and synthesis

5.1 Scientific background

Based on a textual input, a text-to-speech (TTS) system produces a speech signal that corresponds to a vocalization of the given text [All76; Tay09]. Classically, this process can be decomposed in two steps. The first one realizes a sequence of linguistic treatments on the input text, especially syntactical, phonological and prosodic analysis. These treatments give as output a phoneme sequence enriched by prosodic tags. The second step is then the signal generation from this symbolic information.

In this framework, two concurrent methodological approaches are opposed: corpus-based speech synthesis [Bre92; Dut97], and statistical parametric approach, mainly represented by the HMM-based TTS system called HTS [MTKI96; TZ02; ZTB09]. Corpus-based speech synthesis consists in the juxtaposition of speech segments chosen in a very large speech database in order to obtain the best possible speech quality. On the other hand, HTS, which is more recent, consists in modeling the speech signal by using stochastic models whose parameters are estimated *a priori* on a training corpus. These models are then used in a generative way so as to create a synthetic speech signal from a given parametric description.

Corpus-based speech synthesis is a reference since at least a decade. Examples of systems using this technique are ATR Chatr [BT94], CMU Festival [TBC98a], Microsoft Whistler [HAA+96], IBM ViaVoice Text-To-Speech [PBE+06], AT&T Natural Voices [JS02], Loquendo TTS [QDMS01], Microsoft text-to-speech voices, ATR XIMERA [KTN+04], Acapela Voice, Voxygen [Vox13] and IrcamTTS [BSHR05]. This technique relies on systems conceived in the 80s/90s. In particular, diphone speech synthesis has shown to produce results of variable quality but generally with high intelligibility. Corpus-based synthesis can be viewed as an extension of this approach, allowing multiple instances of the units and also variable length units to be considered during a selection step. The problem then turns into finding the optimal acoustic unit sequence. This selection is generally done via a dynamic programming approach such as the Viterbi algorithm. In particular, common implementations of this algorithm try to minimize the audible distortions at junctions between units as well as distances to prosodic and phonological targets [Don98; TBC98b; BRBd02].

Restituted timber quality, which is judged very near to natural, is the main reason of corpus-based speech synthesis success. Another reason is certainly the overall good intelligibility of the synthesized utterances [MA96]. Nevertheless, the main limitation is the lack of expressiveness.

Generally, synthesized voices only have a neutral melody without any controlled affect, emotion, intention or style [Sch01; RSHM09; SCK06]. This is mainly a consequence of the low expressiveness in recorded speech corpora, whose style is often constrained to read speech.

However, expressiveness is an essential component in oral communication. It regroups different speaker and context dependent elements from different abstraction levels which all together enable to highlight an emotion, an intention or a particular speaking style [LM11]. Acoustically, fundamental frequency, intensity and durations of some signal segments are judged to be decisive elements [IAML04; Abe95; Bla07; GR94; IMK+04]. Phonologically, phenomena like phoneme elisions (notably *schwas* in French) or disfluences (e.g., hesitations, repetitions, false starts, etc.) mark different emotional states. At lexical, sentential and more abstract levels, other elements such as the choice of words, syntactic structures, punctuation marks or logical connectors are also important.

The state of the art is presented in [Eri05; Sch09; GP13]. These articles state that current systems have important lacks concerning expressiveness. Moreover, they clearly show the need for expressiveness description languages and for more flexibility in TTS systems, especially in corpus-based systems.

Indeed, controlling expressiveness in speech synthesis requires high level languages to precisely and intuitively describe expressiveness that must be conveyed by an utterance. Some work exists, notably concerning corpus annotation [DGWS06], but for the moment, no language is sufficient to build up a complete editorial chain. This point constitutes an obstacle towards automatic or semi-automatic creation of high-quality spoken content.

The amount of work on the integration of expressiveness into TTS systems is in constant augmentation these last years. Most speech synthesis methods have been subject to extension attempts. In particular, we can cite the diphone approach [BNS02], the corpus-based approach [EAB+04; CRK07], or even the parametric approach [WHLW06; TYMK07]. Adding to this, several languages have been used: notably Spanish [ISA07], Polish [DGWS06], Japanese [WHLW06; TYMK07], English [SCK06], and French [AVAR06; LFV+11]. On our side, current activities in speech synthesis are conducted on French and English. Although other languages could be added to our system, there is currently no real scientific interest, unless a multilingual environment is required.

Beyond speech synthesis, some problems implied by the human expressiveness can also be found in other domains, but generally with an opposed point of view. In speaker processing and automatic speech recognition (ASR), acoustic models try to represent the speech signal spectrum so as to deduce a footprint or to erase specificities and move towards a generic model [SNH03; SFK+05]. In ASR again, the problem of word pronunciations is also important, especially when facing out-of-vocabulary words, i.e.words neither part of the training data nor of hand-crafted phonetized lexicons. Grapheme-to-phoneme converters are then needed to automatically associate one or several phonetizations to these words [Béc01; BN08; IFJ11]. These tools are also used in TTS, needs in TTS and ASR are different. In ASR, the recall over generated pronunciations is maximized, that is the objective is to cover all possible pronunciations of a word to make sure that it will be recognized correctly. At the opposite in TTS, the precision is favored since only one phonetization will be uttered by the system in the end. Thus, extra work on pronunciation scoring and selection is necessary in TTS to improve generic grapheme-to-phoneme models. Other work aims at modeling disfluencies, i.e.errors within the elocution of a sentence, in order to help an ASR system to deal with these irregularities [Shr94; SS96]. By extension, these models are useful to clean a manual or automatic transcription, and make it closer to written text conventions [LSS+06]. Although all these studies share common traits with the expressive speech synthesis problem, they all try to characterize the effects of expressiveness to get rid of

them, and not the other way around. Hence, synthesizing expressive speech requires to extend existing disfluency models to fit a generative process. Finally, emotion detection is also a subject of interest. In [LAVD11], the authors are interested in emotion recognition from linguistic cues while the authors of [SMLR05] propose models mixing acoustic and linguistic features to detect emotions in speech signals. In the case of expressive speech synthesis, dependencies highlighted by these works would have to be reversed in order to predict acoustic features based on given input classes of expressiveness.

In this context, the scientific goal of the team in speech processing is to take into account expressiveness in speech synthesis systems. This objective leads us to the research topics detailed in the rest of the document.

5.2 Challenges

Concerning expressive speech, four challenges can be drawn. Step by step, these challenges split up the entanglement of expressiveness and speech.

The first challenge is, as for the other modalities, the characterization of the expressiveness and variability in human speech, that is the definition of expressiveness in terms of features from all abstraction levels implied in speech production, namely, the signal, prosody, phonology, linguistics, etc. Furthermore, unifying those features through a comprehensive description language is also part of this challenge. This challenge is part of the general challenge C1.

The second challenge is the discovery of dependencies between characteristics of expressive speech, and the construction of models for the recognition and generation of expressiveness in speech. As the development of TTS applications is currently limited by the lack of control of expressiveness, these models should provide fine control facilities while synthesizing speech. As such, they are expected to be highly adaptive and to easily cope with the various natures of speech and language features. This challenge spans over the general challenges C1 and C2.

The third challenge concerns the production of expressive speech resources from scratch. The current common process consists in building a corpus of textual sentences, recording a speaker uttering these sentences and finally annotating the recorded speech. This process is very controlled to guarantee the consistency of speech all over the corpus. We want to extend this protocol to integrate variations due to expressiveness while keeping a consistent view of recordings. To achieve this goal, it is necessary to develop discriminative models able to differentiate speech signals sharing a same lexical content but carrying different expressive signatures, and to annotate them as such. This obviously relies on the definition of similarities and distances between expressive speech signals. This challenge is part of the general challenges C3 and C4.

As an extension, the fourth challenge consists in allowing speech synthesis to rely on open uncontrolled speech databases, especially on massive heterogeneous data from the Internet. In the manner of an information retrieval problem, the key point here is to be able to filter and retrieve interesting speech segments based on comprehensive characterizations by measuring the relevance and consistency of segments according to a target synthesis task. Dealing with heterogeneity and large size of data are two underlying problems of this challenge. As previously, this challenge deals with the general challenges C3 and C4.

5.3 Research topics

As underlined in the previous sections, expressiveness influences an utterance at several description levels. Given a text to synthesize, taking into account expressiveness consists in (i) being able to predict parameters of the different description levels according to a given expressiveness,

and (ii) being able to integrate these parameters in the speech generation process (for example, in the unit selection process). Hence, the main emphasis in speech processing will be put on expressiveness characterization and representation in speech (research focus RF2 of the team), knowledge extraction from expressive data (RF3), generation of expressive speech (RF1 and RF4), and evaluation through realistic use cases (RF5). The higher level of information, devoted to the study of linguistic phenomena, are mutualized with the *text mining axis*. Moreover, the studied language will mainly be French because expertise is necessary to correctly understand variations implied by expressiveness. Nonetheless, English will be also considered since it eases comparison with related work.

Expressiveness characterization and representation: Characterizing the phenomena under study is the first topic as this is also the first challenge of the team. Notably, we propose to study separately elocution styles and affective speech as two sub-domains of expressiveness. As said previously, speech involves a large number of annotations ranging from wide textual structures to narrow speech segments features. Adequate representations and models have to be developed to represent all these levels in a coherent manner. For example, no clear set of expressiveness tags exists to describe the expressive content of a sentence. In the same way, models to predict expressive events, such as particular phonological events (e.g., phoneme substitution), or even prosodic modifications have to be developed. This research topic is shared accross gesture, speech and text processing and will benefit from all advances achieved within the second research focus (multi-level representations).

Knowledge extraction from expressive data: Due to a particular accent, linguistic habits or a particular emotional state, a given speaker will use specific pronounciations, accentuation strategies or rhythms. A possible methodology to mimic these expressive speech peculiarities is to start with well-known generic models and to adapt them according to an application-specific set of constraints. Fundamentally, this approach requires to extract knowledge about how speech and text features interact with expressiveness in order to build these constraints. The extensive use of machine learning techniques is a favored strategy to address this topic. In particular, applying language modeling techniques (for phonology or prosody) on recorded and annotated speech corpora and the development of unsupervised adaptation methods are a clear short term objective. This framework is applicable to tackle intra-speaker as well as inter-speaker variations. This research line will then enable to build speaker or task dependent models of expressive characteristics (prosody, phonology, etc.) and to generate instructions for the speech generation step. This stragegy is also interesting as it is independent of the speech synthesis engine nature and can be used either with unit selection or with a parametric system such as HTS. This topic is part of the third research focus of the team, on knowledge extraction, and it will particularly interact with gesture and text modalities as modeling tools working with both signal and text are required to properly process the spoken language.

Expressive speech generation: Considering the work done within the other topics, expressive speech generation becomes conceivable and thus can be split into the study of expressive phonology (phonetization and disfluencies), expressive prosody (pitch, duration, rythm, pauses), and extension of speech generation algorithms (shortest path finding in graphs of speech units, extension of statistical parametric approaches). These topics can be studied in parallel. Moreover, phonological and prosodic modeling problems are in direct link with machine learning approaches developed within the knowledge extraction topic. This point of view has the advantage to authorize the use

of classical objective evaluation methodologies to conduct a large part of the evaluations. On the contrary, the speech generation algorithmic part and the complete synthesis process need to be perceptually evaluated. This research topic is part of the fourth research focus on generation and, as expressive speech requires high-quality speech signals, of the first research focus dealing with data acquisition.

Use cases and evaluation: Like in any generative framework and directly in line with the team's fifth research focus (use cases and evaluation), it is crucial to evaluate the generated data in concrete use cases, notably by conducting objective and subjective evaluations. Actually, inspite of numerous potential applications, expressive speech use cases are not well developed, mainly because the quality of expressive control is poor in nowadays systems. Consequently, realistic and innovative use cases have to be created as well as adapted evaluation methodologies. While allowing the evaluation of speech synthesis systems, such use cases are also valuable to assess expressiveness annotation techniques, for instance emotion classification techniques.

In all described topics, the question of expressiveness across languages is underlying. Even if this question is not considered as a central problematic, we will try to develop generic methods applicable to several languages. A first step towards this goal is to work on both French and English.

5.4 Application areas

Expressiveness tends to make users accept TTS outputs by producing less impersonal speech. Thus, it plays a fundamental role in a large number of concrete applications. Among all applications, we can mention:

- High-quality audiobook generation;
- Online learning and in particular autonomous language learning;
- Device personalization for disabled people, for whom expressive voice creation is an important need;
- Video games.

More precisely, and to generalize the application potential, the applications of expressive speech rely on the three following applicative functionalities:

- Speaker characterization and voice personalization: models that can be adapted to a speaker thus taking into account its mood, personality or origins. Complete process of voice creation taking into account personalization of voice.
- Linguistic corpus design and corpus creation process: this application domain covers both the design of recording scripts and restriction of audio corpora to address specific tasks.
- High-quality multimedia content generation: this application is really meaningful in the framework of speech synthesis as it needs a fine control of expressiveness in order to keep user's attention.

5.5 Key expected results

- Models of expressiveness in speech both for detection and generation.
- High-quality expressive speech generation with applications to multimedia content generation.
- Massive corpus construction from heterogeneous data: enabling to filter out uninteresting data (for example, in presence of noise, remove tainted parts).
- Algorithms and methodologies to process large collections of heterogeneous data.

6 Expressiveness in textual data

The usage of textual data is dramatically growing: indeed, individuals and organizations communicate and express themselves by using texts, often through Internet both publicly and privately. Textual data may be fully unstructured (a free text) or may be found inside predefined structures (such as web pages, standardized reports, semi-formal models). In the context this research project, textual data may also be considered as transcripts of gesture and speech scenarios. The main research objective is to be able to *identify, characterize, and transfer expressiveness in texts.* In the case of textual data, the definition of expressiveness provided in Section 2 can be made more precise: expressiveness is defined as any variation in text that, while keeping the content semantics, conveys other types of interesting and meaningful information such as style, morphology, and so on. This more specific definition, using a more adapted terminology, is consistent with Figure 3.1 (section 3.1, page 5) where "neutral content" is here meant as the "semantic content" and "expressive content" as "other types of interesting and meaningful information". Expressiveness, as defined above, is quite important for at least two key aspects:

1. Deriving, inferring and extracting implicit information;
2. Characterizing concrete ways for expressing the same semantic content with variations (style, sentiments, etc.)

We consider that for achieving the main research objective, the text axis needs to be based on text acquisition, text mining, and knowledge generation. Text acquisition is required to enrich texts for better specifying both the content semantics and any additional, possibly hidden or contextual, information. Text mining is required for finding, possibly targeted, information within one or several texts. Finally, knowledge generation is required for structuring content semantics and meaning of other types of information for further usage. This further usage generically refers to design and implement computer-based systems facilitating several activities performed by individuals. For instance, (i) making easier, quicker and reliable any choice based on textual data (ii) making explicit hidden information conveyed by textual data and, as a consequence, (iii) enabling understanding of individuals' behaviours, ideas and so on, (iv) making computer-based systems more efficient and effective on the base of available textual data, and finally (v) supporting individuals in following what textual data implicitly suggest. Additionally, in the context of this research project, further usages can be pointed for improving systems supporting gesture and speech scenarios, as mentioned at the beginning of this section.

Accordingly, *all* the three topics needs to be studied to implement the definition of expressiveness. This section first describes the scientific background and challenges of text acquisition, text mining, and knowledge generation. Then, detailed research topics and main outcomes are given.

6.1 Scientific background and challenges

We develop in the following the scientific background along with the related challenges and mutual interactions.

Textual data acquisition and filtering: The first step in order to deal with textual data is the acquisition process and filtering. Raw textual data can be automatically or manually obtained, and need some processes like filtering to be mined. One of this process is the task of corpora annotation. Manually annotated corpora are a key resource for natural language processing. They are essential for machine learning techniques and they are also used as references for system evaluations. The question of data reliability is of first importance to assess the quality of manually annotated corpora. The interest for such enriched language resources has reached domains (semantics, pragmatics, affective computing) where the annotation process is highly affected by the coders subjectivity. The reliability of the resulting annotations must be trusted by measures that assess the inter-coders agreement. Currently, the κ-statistic is a prevailing standard but critical work show its limitations [AP08] and alternative measures of reliability have been proposed [Kri04]. We conduct some experimental studies to investigate the factors of influence that should affect reliability estimation. This challenge deals with the general challenge C1.

Text mining: Due to the explosion of available textual data, text mining and information extraction (IE) from texts have become important topics in recent years. Text mining is particularly adapted to identify expressiveness in textual data. For instance, tasks like sentiment analysis or opinion mining allow to identify expressiveness. Several kinds of techniques have been developed to mine textual data. Sequential pattern extraction aims at discovering frequent sub-sequences in large sequence databases. Two important paradigms are proposed to reduce the important number of patterns: using constraints and condensed representations. Constraints allow a user to focus on the most promising knowledge by reducing the number of extracted patterns to those of potential interest. There are now generic approaches to discover patterns and sequential patterns under constraints (e.g., [NLHP98; PHW02; PHW07; Bon04]). Constraint-based pattern mining challenges two major problems in pattern mining: effectiveness and efficiency. Because the set of frequent sequential patterns can be very large, a complementary method is to use condensed representations. Condensed representations, such as closed sequential patterns [YHA03; WH04], have been proposed in order to eliminate redundancy without loss of information.

The main challenge in sequential pattern extraction is to be able to combine constraints and condensed representations as in itemsets paradigm which can be useful in many tasks as to analyze gesture and speech captured data.This challenge spans over the general challenges C1 and C3.

Knowledge generation: Knowledge generation consists in organizing the information which can be manually or automatically extracted from texts and in representing it a compact way. This representation can vary according to the adopted abstraction level. When studying texts as sequences of words, this representation can be referred to as a language model, while knowledge will rather be represented as ontologies when considering more abstract, higher level, views of texts.

Language models aims at deriving and weighted short linguistic rules from texts, typically using statistical approaches, in order to approximate their shallow structure. These rules are useful to compare texts [SC99] or to help applications in choosing the most likely utterances among a large set of candidates. For instance, language models are used in machine translation [MBC+06], paraphrase generation [QBD04] or ASR [RJ93] to ensure against ungrammatical output texts. The most widely spread language modeling technique is the n-gram approach [Jel76], but major

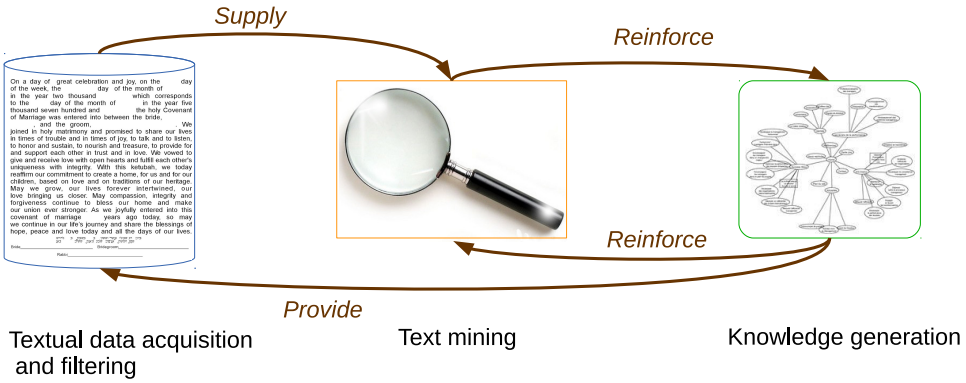

Figure 2: Possible interactions within the text axis.

advances have been achieved recently, leading to outperform this venerable approach. Especially, methods based on neural networks exhibit very good performances [SG02; BDVJ03; MKB+10; MDK+11; Mik12]. As these models are still new, they need to be further studied and extended. In the scope of the team EXPRESSION, these models should be used to model expressive texts.

Second, ontologies are a tool enabling explicit and precise representation of information and knowledge about concepts and relationships hidden in available texts. Indeed, textual data provide samples of concepts and relationships (such as words 'my car...' as example of a possible concept 'car'), as well as references to concepts and relationships (such as word 'car' as reference to a possible concept 'mean of transport' or just to 'car'). Finding those concepts and relationships is a prerequisite to further enrich earlier ontology versions by adding new artefacts (for instance, new axioms), not (necessarily) provided in available texts. However, as also recently highlighted [Gan13], despite the work performed, there is still the need to understand much better how to bridge the gap between; on the one side, techniques usable for processing and analyzing texts and, on the other side, information for filling in ontology content (basically concepts, relationships, axioms). Understanding foundations leads to automation improvement and therefore reduction of the required human effort for extracting valuable ontologies.

Hence, a major challenge for the team is to propose solutions to integrate textual expressive information into these knowledge representations. This challenge is part of the general challenge C3.

Interaction between acquisition, mining, and knowledge generation: A key important point concerns interactions between the three aforementioned domains. Interactions are required to improve, by reusing methods and techniques proper to each of them, solutions proposed for solving indiviual challenges. Therefore, interactions should be elicited, planned and coordinated as part of the research activity. Figure 2 shows possible interactions (arrows in the figure) between the text processing domains of interest. For instance, some concrete examples of these interactions are:

- Text mining can be helpful for early steps of knowledge generation by highlighting interesting segments and phenonema to be studied in texts; similarly, generated knowledge can be helpful for constraining text mining;

- Ontology learning (a task of knowledge generation) Knowledge representations can be improved by using mined patterns in order to generate semantic relations, concepts and in-

stances within an ontology or a statistical model;

- Knowledge generation is needed to characterize the content semantics; annotations, resulting from text acquisition, are required to introduce additional information for further characterizing variations according to any knowledge generation process.

Bringing interoperability between proposed methods and developing such interactions is one of the challenges of the text research axis and contributes to the general challenge C2.

6.2 Research topics

Many research topics exist in order to deal with expression in textual data. In particular, we focus on the four following topics.

Corpus annotation: Currently, annotated corpora are needed to evaluate a system of multi-document summarization, which uses a similarity measure between sentences as a preliminary task. We are conducting multi-coders annotations for which subjectivity of the coders is high: measuring the similarity of pairs of sentences and measuring the informative power of sentences. These experiments extend and increase our investigations of annotated data reliability. This research is conducted with the LI team of University of Orleans-Tours. It is currently a work in progress. We plan to extend all these experiments with simulated synthetic data to characterize precisely the relations between absolute reliability measures and expected confidence in the reference annotation. This work is part of the research focuses RF1 (data acquisition) and RF5 (use cases and evaluation) of the team.

Sequential pattern mining: We focus on the use of sequential patterns with textual documents (RF3 of the team). Sequential patterns can be useful in many tasks like log analysis, recommendation, event detection, stylistic analysis. In former work, we have focused on the use of sequential patterns on biomedical tasks [BCC+12] and geographical data [AB14]. We also work on approaches using sequential patterns in order to identify expressiveness as appositive qualifying phrases [BCCC12]. All these work are based on algorithms allowing to extract sequential patterns with multiple kinds of constraints like numeric or linguistic constraints. However, many challenges still exist in sequential patterns extraction in order to 1) reduce the number of extracted patterns and 2) to improve the quality of extracted patterns. We particularly focus on this problematic by working on the development of an algorithm of sequential patterns extraction under constraints having an exact condensed representation. Other works are still in progress like using techniques of formal concept analysis and recovery measures to select relevant patterns, or also using clustering. Another research that we plan to focus on is the adaptation of graph mining algorithms, like the well known *GSPAN* [YH02] and *GASTON* [05] algorithms, to the specific case of textual data. Indeed, existing algorithms in the literature currently deal with generic graphs and are not adapted to textual data. We particularly focus on oriented graphs in order to mine dependency trees allowing to have more complex and informative patterns as sequential patterns give. Finally, we focus on named entity recognition by combination of linguistic and statistical (SVM) methods. The same approaches were tested to characterize expressiveness of parenthetical segments (that is texts between parentheses). This research topic contributes to the general research focus RF3 (knowledge extraction).

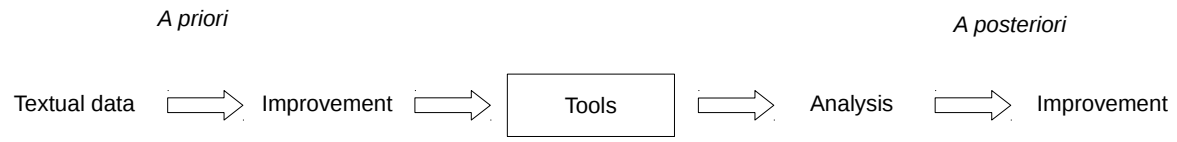

Figure 3: Improve the quality of textual data for the purpose of ontology building

Language modeling with multilevel information: Textual utterances can be represented at many different but interconnected levels, for instance through morphosyntactic, syntactic, semantic, and lexical information. Such a rich and diverse representation is particularly needed to study expressiveness. Hence, processing expressive texts for a given application often requires not only to figure out a sequence of words but more effectively to represent and combine many features. An important research topic towards the modeling of expressive texts is thus the development of multilevel information language models, and particularly statistical models. Recent language modeling paradigms based on neural networks will especially be studied, as work on these models has already been engaged previously in the context of ASR [LM12]. On the one hand, this research topic seeks to enable automatic text generation with expressiveness constraints, the latter coming either from linguists of from various automatically extracted linguistic patterns. This research topic shares some issues with the field of machine translation and paraphrase generation. On the other hand, language models including features of expressiveness are also a step towards similarity measures between expressive texts. As such, this research topic contributes to many global research focuses, namely the research focuses RF2, RF3, RF4 and RF5 (multi-level representations, knowledge extraction, generation, and use cases and evaluation).

Ontology building and improvement: To bridge the gap between (i) techniques usable for processing and analyzing texts and (ii) information for filling in ontology contents (basically: concepts, relations and axioms), we have undertaken an approach integrating two distinct perspectives:

- improving the quality of textual data for the purpose of ontology building (or learning) and
- improving generated ontologies by detecting and possibly solving known problems and defects in them.

While the first perspective is clearly related to incoming data and their quality, the last point is related to the general problem of "ontology quality and ontology quality improvement". Within the approach, two main ways can be followed for generating an ontology. According to the first way, ontology building is manually performed by human starting from available high quality texts. However, a predefined meta-ontology framework constraints humans whenever they introduce concepts and relationships [HBO12; OBHM12]. According to the second way, existing tools for automatically or semi-automatically building ontologies from texts are analyzed, compared and selected according to their "performance" within a test-bed platform [GHBK13; GBHK12]; then, possible problems and improvements of resulting ontologies are investigated (some preliminary results can be found in [CHB13]). An additional third way has been experimented in the case of textual data used for building ontologies concerning human competencies. This third way combines both a meta-ontological framework (specifically designed for the context of human competencies and required to precisely define what is expected as ontology content) and tools for extracting information and knowledge from texts (required to identify individuals, competencies and other contents, as available in texts concerning activities performed by those individuals, e.g. CV and other

textual resources) [BDHS11]. Performed experiments put in evidence several weak points, introducing errors in resulting ontologies. Two weak points can be mentioned: the general applicability to any type of text and identification of known relationships. These two points will be subjected to further research in the context of the EXPRESSION research project. This work on ontologies takes place in the research focuses RF3 (knowledge extraction) and RF5 (use cases and evaluation).

6.3 Application areas

Significant application domains are given below.

- Under-resourced language analysis will be made possible for instance by developping new tools (like POS Tagger, syntactic parser) for unusual languages (as Latin or Sanskrit), based on sequential pattern extraction.
- Video games, plagiarism detection, recommender system are instances of applications where extraction and transfer of different expressive forms within textual data (like language registry or state of mind) using patterns or rules will be very useful.
- Opinion mining and sentiment analysis will benefit from new corpora of French emotional norms, i.e. dictionaries which give the polarity of each entry. Building such resources is very expensive and automatic processes have to be tested to extend manually built norms [VB11].
- Human-machine dialogue systems (potentially including TTS) will be improved by integrating expressive models and features, enabling for instance text modulation in order to fit users' profiles.
- Information retrieval and automatic summarization systems are applications where semi-automatically built ontologies will provide a better understanding of texts.

6.4 Key expected results

- Produce a new efficient algorithm to extract sequential patterns with condensed representation and constraints, which can be more efficient to deal with important data sets.
- Develop a prototype to extract and transfer different forms of expressive specificities.
- Build a model of pattern with important level of granularity for opinion mining.
- Integrate expressive features and multiple information levels in language models and develop expressive language models.
- Improve the quality of automatically built ontologies from texts.

7 Main expected outcomes and synergies between modalities

The synergies between the three modalities and the main expected outcomes are mainly relying on methodology, framework and platform factoring as depicted in Figure 4 and detailed below. Furthermore, the whole is expected to be more than the sum of its parts: Namely the knowledge we build from the study of one media is expected to be useful to the characterization of the phenomenology that we may face in other modalities.

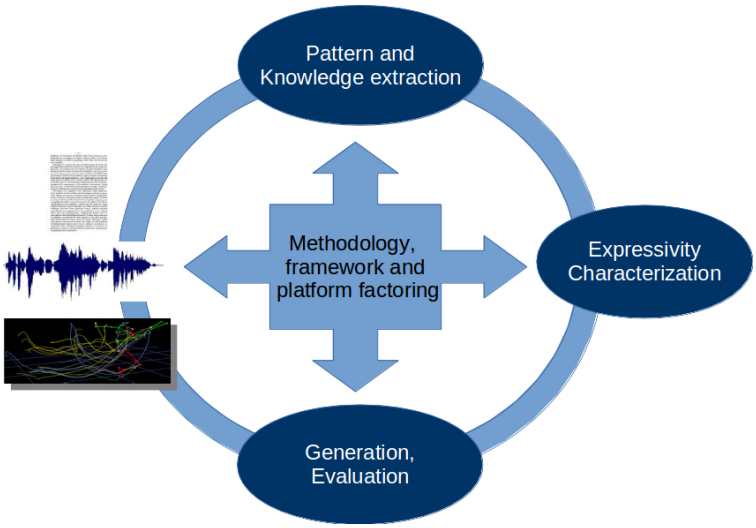

Figure 4: Synergies and main expected outcomes.

Characterization of expressiveness for the three modalities gesture, text, and speech: One of the main expected results is somehow a sufficient qualitative and quantitative understanding of what makes expressiveness in these modalities in order to reproduce or mimic it. Comparing the way expressiveness is "encoded" into these three medias will be a very valuable outcome allowing some cross-domain applications and leading to open original or unexpected research directions.

Methodologies factoring: All the three domains addressed by this research project share common methods. For example, gesture synthesis and speech synthesis rely on identical key elements such as pattern selection and concatenation. We believe that the study of sequential patterns at numerical and symbolic levels will bring relevant highlights for the three considered medias. We think that other methods, in particular data driven methodologies and associated machine learning approaches, ontology modeling, evaluation protocols (quantitative and perceptual), etc., may be translated from one domain to another and thus provide benefits to the whole community.

Integration of new expressive descriptors and models in the generation processes: By understanding how expressiveness works, we will be able to propose a set of features to describe it, and develop models to explicitly control the generated output, whether it is gesture or speech, or even text if we are able to consider paraphrasing under expressiveness constraints.

Development of common frameworks and platforms: Both for corpora annotation and for evaluation purposes, the development of platforms is necessary. Considering the nature of manipulated data, we think possible to develop common frameworks (for multi-level data representation, indexing and retrieval, and modeling) or platforms (e.g., for subjective evaluation).

8 Positioning

In this section, the Expression team is positioned into its national and international communities. Then, the originality of the Expression team within IRISA lab. is also given before

detailing the strategy to make the team visible in its various related communities.

8.1 National and international related teams and laboratories

We briefly report the main national and international teams whose activities are highly correlated to the research program proposed by the EXPRESSION team. A larger list of potentially competing or complementary teams (although less focused on the research axis targeted by EXPRESSION) is proposed in the appendix. However, we stress that the proposed list of teams and lab. is not exhaustive, as the fields related to our research are varied and numerous.

Expressive gesture

For the past decades, there has been a growing interest in Embodied Conversational Agents (ECAs) to be used as a new type of human-machine interface. These agents are virtual characters with human-like appearance and skills, which can communicate with a user, combining both verbal and non verbal behaviors, with information flows including spoken language, facial movements (facial expression, gaze direction, etc.), hand gestures, and body movements. Several teams have conducted studies to represent and model emotional expressive agents. The LTCI Lab. (Télécom ParisTech) has been developing for several years an interactive embodied conversational agent (ECA) platform in order to make human computer interactions more believable, taking into account social and emotional behaviors. In this line of research, the Sociable Agents Group (Center of Excellence Cognitive Interaction Technology – Bielefeld University – Faculty of Technology) studies the coordination and adaptation of the virtual agent's behavior with the user's one, according to social attitudes and contexts. Using 3D virtual agents in interactive situations, the group Cognition, Perception, and Usages of the *Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur* (LIMSI) studies the role of human bodily postures in interactive situations to characterize expressions of emotions. The team IHSEV (virtual human axis) (Lab-STICC) also focuses on interaction between human and artificial systems. Other groups conduct researches in the domain of affective computing at MIT (http://affect.media.mit.edu/), combining computer science with other domains (psychology, cognitive science, neuroscience, sociology, etc.), or focus on autonomous synthetic characters, with interests in affective computing, interactive narrative and human-robot interaction (Ruth Aylett, Heriot-Watt University, Edinburgh, UK). Most of the time, the former approaches use explicit behavioral models, whether based on cognitive theories, communication functions, or linguistic models, and are especially interested in analyzing the interaction between real and virtual human for different goals and applications (dealing for example with rehabilitation, social training, or performative arts). We are not directly interested by explicit behavioral approaches, but we study more deeply gesture models, without dissociating analysis from synthesis and coupling low level to high level models. Hence, the main goal is to extract knowledge from data, both at low levels, using statistical and machine learning models, and at high levels, linking continuous signals to annotated or linguistic patterns.

Recent research in the use of 3D virtual human characters supports sign linguistics research, providing a potential target for symbolic translation of sign languages, animation of sign language using avatars, and usability evaluation of practical translation and animation systems. Three successful international workshops, held in Berlin, Dundee, and Chicago, brought together researchers in this emerging field. Among the involved teams we can cite: Athena RC, Institute Language and Speech Processing (Greece); University of Hamburg, Institute of German Sign Language and Communication of the Deaf (Germany); University of East Anglia, School of Computing Sciences

(UK); Rochester Institute of Technology (RIT), City University of New York (US); LIMSI-CNRS (Sign Language Processing Team, France); IRIT (Image processing and sign language research, France); DePaul University, Chicago (US); Embodied Agents research group at the DFKI (German Research Center for Artificial Intelligence, Saarbruck, Germany).

In computer animation domain, many research teams are also interested by analysis / synthesis methods to automatically animate virtual characters. These approaches may consider style in motion generation, thus leading to more realistic motion, but most of the time there is no linguistic dimension in the models. However, there is a growing interest towards this type of concern, as in the Multimodal Computing and Interaction group in Saarland University who proposes automatic multimodal annotations, and thus takes into account spatial and temporal variations present in semantically related motions.

Other connected research domains, especially those dealing with computer music and performing arts have progressively integrated the expressive gestural component in the design of interactive multimodal human-computer interfaces. For example the Input Devices and Music Interaction Lab. (IDMIL) at McGill University, with whom we have worked since 2006, deals with research related to the design of musical instruments and interfaces for musical expression, movement data collection and analysis, sensor development, and gestural control. Also connected to expressive gesture, the IMTR team at Ircam which conducts research on gesture analysis and modeling for real-time musical interactions, and the InfoMus Lab (CASA Paganini) at University of Genova which carries out research on computational models of non-verbal expressive and social behaviors.

Expressive speech

Although a large number of research groups are working on text-to-speech synthesis, less are focused on expressivity in speech. Considering the quite high quality of synthesized speech, the main interest for the speech community is now to understand what makes expressivity and how to reproduce it.

Several teams work on the generation of prosody from phonological and linguistic content, specifically targeted to the speech synthesis domain, such as the *Laboratoire de Linguistique Formelle* (LLF) with the work of Elisabeth Delais-Roussarie on prosodic interfaces, prosodic phonology. On this subject, we share common interests in understanding how speech synthesis can be improved toward the generation of expressive contents with a high control on prosody. The *Laboratoire de Phonétique et Phonologie* (LPP) also works on related domains with a focus on corpora construction for speech recognition and synthesis with natural language processing approaches. In particular, these teams do not consider the whole processing chain from text to acoustic for expressive speech generation.

Other teams, such as the Sound Analysis and Synthesis group at IRCAM, are more interested in the acoustic processing part. Thus, this team works on sound analysis, transformation, and synthesis of sound signals, mainly for musicians. However, their work can be applied to related domain such as video games and virtual reality. They also work on expressive speech synthesis with recent publications on this subject, using classical text-to-speech frameworks like HTS.

Moreover, at LORIA, the PAROLE team works on spoken communication and covers a wide spectrum of domains, notably, speech recognition and synthesis, speech analysis or articulatory modeling. One specificity of this team is their know-how in the field of articulatory speech synthesis as well as in multimodal speech processing. At the LIG, in Grenoble, the team GETALP

principally works on automatic speech transcription and translation, under-resourced languages processing, speech and interactions in noisy environments analysis and processing, social affects modeling. For this last point, the work conducted by Véronique Aubergé on the prosody of attitudes or emotions is of particular interest. Complementary to these topics, the audio and acoustic group of the LIMSI also works on expressive prosody trying to understand what parameters are used to encode expressivity in speech. Moreover, speech synthesis is also addressed by this team, considering mainly the acoustic signal processing, modeling and perception for the expressive aspects of voice.

At the international level, several teams work either directly on expressive TTS or on related domains such as affective computing. Using a corpus-based approach, AT&T Research, New Jersey, USA, produces high-quality voices for English. Moreover, the Centre for Speech Technology Research (CSTR), Edinburgh, UK, notably works on speech synthesis mainly using statistical approaches. Recently, they engaged a work on synthesis using deep architectures for statistical synthesis.

Concerning emotions representation, we can mention the Signal Analysis and Interpretation Lab, University of Southern California, Los Angeles, USA and the Technical University of Madrid, Department of Electronic Engineering who work on the subject. In particular, the former is interested in the creation of emotion profiles and the perception of affective/emotional states while the latter focuses on emotional speech synthesis.

Finally, another important research center is the Toshiba Cambridge Research Laboratory, UK who is working on expressive synthesized speech using HTS and also deep neural networks. They conduct research on acoustic factorisation in order to control separately speaker characteristics and language.

Considering this academic research environment, some potential partners can be identified either sharing common topics, such as CSTR, LORIA, TCTS Lab, or IRCAM, or working on complementary fields. For instance, collaborating with LLF or the LIG on the fine control of prosody adapted to expressivity or with the LPP on higher levels of the speech synthesis process, as pronunciation modeling or more generally linguistic content adaptation, is to be considered.

The industrial environment is also important with some major companies like Acapela, Ivona, Loquendo, and Voxygen. These companies work on TTS including expressive and emotional aspects. On a methodological point of view, the current trend is to use multi-expressive voices by the aggregation of several expressive corpora, each dedicated to one particular kind of expressivity. In our opinion, such approaches mainly rely on a costly annotation of dedicated corpora which we can try to relax.

In this global picture, speech activities of the Expression team are original as they are positioned at the frontier between signal processing and formal linguistics. Second, few teams have a strong expertise on French, both with unit selection and parametric approaches. Finally, the synergy with gesture and text modalities is unique to our knowledge. This should clearly push forward research by introducing new solutions and tools to speech-related problems, and innovation by enabling new multimodal applications.

Expressive text

According to Section 6, expressiveness in textual data is a complex multifaceted process, jointly employing theories, methods and techniques developed in the three scientific pillars: text mining, knowledge modeling and text data acquisition. Therefore, for positioning, both nationally and internationally, the research activity undertaken within the "text" axis, only research teams adopting a similar joint research strategy have been shortly introduced in the remainder. Other

remarkable teams are simply mentioned.

Regarding the national context, the team *Représentation des Connaissances et Langage Naturel* (RCLN) of the *Laboratoire d'Informatique de Paris-Nord* (LIPN) mainly deals with computational linguistic problems and focuses on key problems like pattern extraction, dynamic semantic annotation, and the speech-syntax integration. These problems are eventually close to problems mentioned in Section 6, driving some of our research activities. At LIMSI, the NLP group works on the development and evaluation of NLP systems, on information extraction from texts (specifically patterns, syntactic parser, etc.), and on specific facets of expressiveness (such as opinion mining and sentiment analysis). Other national teams can also be mentioned even though their research strategies are loosely related to expressiveness (in textual data); they are: TALN team (LINA), MELODI team (IRIT), LaTTiCe, ORPAILLEUR team (LORIA).

In an international perspective, Bristol Centre for Linguistics at University of the West of England is specialized in lexical analysis, analysis of intentional lexical irregularities, linguistic approach for building patterns. The Laboratory for Applied Ontology part of ISTC-CNR, deals with ontological foundations of conceptual modeling, knowledge representation, knowledge engineering, natural language processing, semantic web, and so on. Finally, an important research centre focusing on pattern extraction and data-mining is the Data Mining Research Group at the University of Illinois.

Furthermore, we are observing quite recent initiatives putting in evidence the interest of the joint research strategy founding the scientific base of the "text" axis. Among other, the Language Technologies Institute (LTI)[3] at Carnegie Mellon University is producing a striking effort in this direction.

8.2 Related teams at IRISA

Within the IRISA ecosystem, the proposed project is positioned inside the Media and Interaction Department (MID, Dpt.6). It addresses an area of research which is complementary to the ones covered by the MIMETIC (realistic virtual human modelization and design), LINKMEDIA (multimedia linked distributed data information processing and retrieval) and INTUIDOC (analysis, recognition, interpretation of digitized documents, and man-document interaction) teams, all of them also part of the MID department.

- Compared to MIMETIC which mainly addresses sport movement and does not cover the language dimension conveyed by gesture, nor the expressive quality of gesture, EXPRESSION focuses on the characterization, synthesis and recognition of expressive gestures, with original applications that target the processing of sign languages and performing art gestures.

- Compared to LINKMEDIA which focuses on the characterization and retrieval of multimedia linked data within a big data context, EXPRESSION does not address the potentially linked (social) or big nature of speech, text and gesture data. Complementarily, EXPRESSION addresses the generative process of expressive speech, textual, and gestural data, which is not dealt with by LINKMEDIA.

- Compared to INTUIDOC whose activity is centered on the writing communication and the engineering of written documents, mainly exploiting the text and gesture media, the object of study for EXPRESSION is rather the characterization and exploitation of expressiveness in speech, text and gesture data, namely in synthesis, recognition or mining tasks.

[3] `https://lti.cs.cmu.edu/`.

8.3 Strategy to develop visibility and impact

The team will essentially target main conferences and journal for each of the addressed modalities while also publishing in conferences and journals mixing together the different modalities. Top ranked journals will be targeted to disseminate major transverse results, particularly theoretical and methodological results, but also innovative applications. Table 1 highlights the most important places where the team will publish. For each conference and journal, a very brief description is given and the level of importance is given for each modality.

In addition, the team will make available to the research community resources (e.g. corpora, benchmarks), dedicated codes and software to improve its impact. Furthermore, the participation to collaborative research projects at national and international levels will be encouraged as a factor of visibility.

9 Team experience

This section lists our on-going projects and collaborations, as well as the software developed by the team and internally available technical facilities.

9.1 National on-going projects

- Programme Investissements d'Avenir (Usages, services et contenus innovants) : 2012-2014
  - SIGN3D
  - Partners: Mocaplab, Websourd, IRISA
  - Subject: Editing motion in French sign language using high definition gesture databases
  - Summary: The Sign3D project aims at creating a range of innovative tools for the recording and the editing of captured motion of French Sign Language (LSF) content. The challenge is to design a complete workflow from the movement capture (including body and hand movements, facial expressions and gaze direction) to the restitution using concatenate synthesis applied on a 3D virtual signer.
  - `http://sign3d.websourd.org/`
  - UBS Participants: Ludovic Hamon (post-doctorate), Sylvie Gibet
  - 160 k€ + Pôle Images & Réseaux
- ANR project CONTINT: 2012-2016 (Coordinator: Pierre de Loor, LabSTICC)
  - INGREDIBLE
  - Partners: LabSTICC, LIMSI-CNRS, IRISA, Virtualys, Final users (DEREZO, Brest; STAPS lab., Orsay)
  - Subject: Analysis and synthesis of expressive gestures in interactive theatrical scenarios by machine learning methods.
  - Summary: 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.
  - UBS Participants: Pamela Carreño, Sylvie Gibet, Pierre-François Marteau, Ludovic Hamon, Caroline Larboulette

Name	Association	Gesture	Speech	Text
Journals				
Computer speech and language	ISCA	+	+++	++
Pattern analysis and applications	Springer	++	+	++
Pattern recognition	Elsevier	+++		
Signal processing letters	IEEE	++	+++	
Transactions on affective computing	IEEE	++	++	++
Transactions on audio, speech, language processing	IEEE/ACM		+++	++
Transactions on neural networks and learning systems	IEEE	+++	++	++
Transactions on pattern analysis and machine intelligence	IEEE	+++	++	++
Universal Access in the Information Society	Springer	++	+	+
Computational Linguistics	MIT Press/ACL	+	++	+++
ACM Transactions on Interactive Intelligent Systems	ACM	++	++	++
Computer Animation and Virtual Worlds	Wiley	+++	+	
Journal on Multimodal User Interfaces	Springer	++	++	++
Transactions on Knowledge and Data Engineering	IEEE			++
Conferences and workshops				
ACL: Meeting of the Association for Computational Linguistics	ACL		++	+++
ACM Symposium on Applied Perception	ACM	++	++	
AMFG Analysis and Modeling of Faces and Gestures	IEEE	+++	+	
CASA: Computer Animation and Social Agents	?	+++	+	++
EMNLP: Conference on Empirical Methods in Natural Language Processing	ACL		++	+++
Expressive	ACM	++	++	++
Humanoids	IEEE	+++	+	
ICASSP: International Conference on Acoustics, Speech and Signal Processing	IEEE	+	+++	
ICDM: International Conference on Data Mining	IEEE			++
ICMI: International Conference on Multimodal Interaction	ACM	++	++	++
ISWC: International Semantic Web Conference	Springer			+++
IVA: Intelligent Virtual Agent: mixing modalities (speech, gesture, text)	Springer	+++	++	++
International Society for Gesture Studies	ISGS	+++		
Interspeech	ISCA		+++	+
LREC: Language Resources and Evaluation Conference	ELRA	++	++	++
MIG: Motion in Games	Springer	++		
SIGKDD: International Conference on Knowledge discovery and data mining	ACM			++
SSW: Speech synthesis workshop	ISCA		+++	
TSD: Text speech and dialogue	ISCA		++	++

Table 1: List of the main journals and conferences for the team Expression.

  - 160 k€ + Pôle Images & Réseaux
  - PhD : Pamela Carreño (à partir de 01/10/2012)

- ANR project Phorevox: 2012-2014 (Coordinator: Damien Lolive, IRISA)
  - Partners: IRISA, LLF, CREAD, Zeugmo, Voxygen
  - Subject: Creation of dynamically generated contents relying on vocal synthesis to support the development of written skills
  - Summary: Technological solutions for language learning, and specifically for french, are almost restrained to exercises relying on writing skills (QCM, word input or pictograms). We believe that oral interaction can help students or pupils improve their writing skills. Educational exercises, which cover several language acquisition phases are proposed, namely from phonological acquisition to more traditional exercises such as word or sentence dictations and word segmentation. All of them target the acquisition of skills necessary to free text production. To achieve this goal, five groundbreaking research areas are explored: build high quality synthesized voices, define exercices usable to work on specific linguistic skills, propose a exercise vocalization tool on a web platform enable to gather students' answers, conceive a skills profiling tool and propose an automatic exercise generation mechanism using the student profile.
  - UR1 Participants: Nelly Barbot, Jonathan Chevelu (post-doctorate), Damien Lolive
  - 140k€ + Pôle Images & Réseaux

- ANR project Hybride: 2011-2015 (Coordinator: Yannick Toussaint, INRIA/LORIA)
  - Subject: 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).
  - Participants: INRIA/LORIA, GREYC, MoDyCo, Inserm
  - IRISA is associated to the GREYC in this project

- CNRS MASTODONS project ANIMITEX: 2013-2015 (Coordinator: Mathieu Roche, Cirad/TETIS)
  - Subject: 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.
  - Participants: LIRMM, TETIS, ICUBE, GREYC, LIUPPA
  - IRISA is associated to the GREYC in this project

9.2 International on-going projects

- Collaboration with IDMIL-CIRMMT, University of McGill, Montréal (Canada):
  - PhD + postdoctorate (2006-2010);
  - PhD (50% McGill, 50% ARED) started in October 2012: Sketched-based gesture interaction for data-driven musical sound control (Lei Chen)

- Collaboration with Gallaudet University, Washington, USA (PhD Kyle Duarte defended in June 2012)

9.3 Industrial cooperation

- Mocap Lab: motion capture (project SIGN3D)
- Websourd: sign language usages and evaluation (project SIGN3D)
- Thales (Optronic, TOSA): gesture recognition for control/command of robotized vehicle (CIFRE funded program)
- Virtualys: integrated Platform (project INGREDIBLE)
- Voxygen: Speech synthesis (ANR project Phorevox)

9.4 Defended PhDs

Guiyao Ke. Mesures de comparabilité pour la construction assistée de corpus comparables bilingues thématiques. informatique. Université de Bretagne Sud, Feb. 2014. French

Thibaut Le Naour. Utilisation des relations spatiales pour l'analyse et l'édition de mouvement informatique. Université de Bretagne Sud, Dec. 2013. French

Sébastien Le Maguer. Évaluation expérimentale d'un système statistique de synthèse de la parole, HTS, pour la langue française. Université de Rennes 1, Jul. 2013. French

Kyle Duarte. Motion Capture and Avatars as Portals for Analyzing the Linguistic Structure of Signed Languages. Université de Bretagne Sud, Jun. 2012. English

Muhammad Fuad M. M. Similarity Search in High-dimensional Spaces with Applications to Time Series Data Mining and Information Retrieval, Université de Bretagne Sud, Feb. 2011. French

Charly Awad. Indexation et Interrogation de Bases de Données de Mouvement pour l'Animation d'Humanoïdes Virtuels. Université de Bretagne Sud, Feb. 2011. English

Larbi Mesbahi. Transformation automatique de la parole : étude des transformations acoustiques. Université de Rennes 1, 2010. French

Alexandre Bouënard. Synthesis of Music Performances: Virtual Character Animation as a Controller of Sound Synthesis. Université de Bretagne Sud, Dec. 2009. English

9.5 Software

Unit selection speech synthesis engine

For research purposes we developed a whole text-to-speech system designed to be flexible. The system, implemented in C++, intensively uses 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 manipulating large speech corpora such as hash tables and pre-selection filters [CBSP00]. 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 engine has been published in [GL14a] and [GL14b].

Library and toolkit Roots

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. In practice, the toolkit 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.

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 [BCLL12]. More recently, ROOTS has been registered through 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].

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.

Software LamSCP

Building a voice for TTS purposes generally relies on a recording script extracted from a huge text corpus and optimized to cover linguistic and phonological events which are supposed to lead to a good voice acoustic quality. As many events are rare, the main difficulty is to assure the presence of all of them while minimizing the speech recording duration. Indeed, a short speech corpus tends to guarantee the quality and homogeneity of the voice and to minimize the recording and post-processing costs. Designing such a rich corpus with a minimal size can be formulated as a Set Covering Problem (SCP). As this problem is NP-hard, one needs the use of heuristic based algorithms to reduce huge corpora, i.e. containing more than one million sentences.

The team has developed a full workflow to solve this problem. An efficient and parallelizable tool has been implemented in C++ to build feature dependent matrices based on an initial sentence set to be reduced. These matrices are sparse and their size can reach up to 50,000 rows and more than one million columns. Several optimization algorithms have been implemented, based on greedy strategies and inner optimization procedures. Moreover, an innovative algorithm called LamSCP has been proposed to solve SCP on huge corpora with multi-representation constraints. LamSCP relies on Lagrangian relaxation properties and gives better results than state-of-the-art

greedy algorithms (returned reduced sets are from 5 to 10 percent shorter). Even more interestingly, this algorithm provides a lower bound to the optimal solution cost for the considered SCP. This lower bound permits to assess the closeness of the calculated solution to the optimal one: it turns out that LamSCP and greedy algorithms give solutions close to optimal [CBBD07]. This study of the SCP and the associated algorithms for speech processing have been published in several conferences [ABB+08; CBBD08] and [BBD12].

Library for editing LSF and concatenative synthesis of gestures

We have developed in the team a whole motion-capture-driven synthesis pipeline dedicated to the editing of gestures in French Sign Language (LSF) and the generation of gestures that can be visualized through a 3D virtual signer. This system is able to produce novel utterances from the corpus data by combining motion chunks that have been previously captured on a real signer, and by using these data to animate a virtual signer. As gestures in signed languages are by essence multichannel, i.e. meaningful information is conveyed by multiple body parts acting in parallel, it follows that a sign editing system manipulates motion segments that are decomposed on these channels over time. The editing system is able to accurately and efficiently retrieve the annotated SL items from the database, and to concatenate the corresponding motion chunks spatially (i.e. along the channels), and temporally within an animation system. This last one thus synchronizes and handles at the same time several modalities involved in signed gestures and produce a continuous flow of skeleton postures.

This development has given rise to the achievement of a set of software libraries described below:

- The library "Sgn_Core" is dedicated to the recording and the processing of one or several movements, and one or several annotations extracted from "fbx", "bvh" and "eaf" files.
- The library "Sgn_Db" is dedicated to the recording, the decomposition, the indexing, and the extraction of all or part of several captured and annotated movements from a heterogeneous database.
- The library "Sgn_Edit" is dedicated to the creation, the editing, and the visualization of a motion produced by recomposing the processed motion chunks.
- The library "Sgn_IHM" is dedicated to the visualization of the motions of one or several joints within a 3D virtual environment.

These libraries have been registered through the Program Protection Agency (*Agence pour la Protection des Programmes*, APP).

Sequential Data Mining under Constraints (SDMC)

SDMC is a web site having a set of text and data mining tools. This web is developed with the support of the Hybride ANR, with different partners: GREYC, LIPN, MODYCO laboratories. Link: `https://sdmc.greyc.fr/`.

9.6 Technical facilities and platforms

Speech recording studio

A main goal of the Expression project consists in developing high quality voice synthesis. A good speech corpus quality relies on a consistent speech flow (i.e., 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. Hence, the team owns a speech studio located at ENSSAT in Lannion. The studio, in its material form, comes along with a software platform developed internally.

The 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. In addition to audio signal, Electro-Gloto Graphs (EGGs) can be captured from the actor. This activity is used to induce the F0 (first formant) trajectory which is the main indicator of the prosody.

Regarding software, recording sessions are orchestrated using a dedicated tool. Mainly, the role of this tool is to prompt actors in the isolation booth to utter speech with various indications (mood, intonation, speed, accent, role, ...). The prompt is presented on a simple interface. Then, sound files are recorded, segmented and linked to the transcription. The whole process is controlled by the operator in real time. The latter can possibly reject (in fact, annotate) a file and prompt the actor again with the discarded sentence in case of mispronunciation, bad audio quality, etc. This software has been developed in C++ and relies on the Windows Audio and Sound API (WASAPI).

Listening test platform

The listening test platform is developed by the team especially to evaluate speech synthesis models. This platform has been developed to propose 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).

Motion captured recording platform

The expressiveness of expert gestures requires accuracy and high definition in the recording of captured motion. We have acquired in the team an expertise in the capture, post-processing, and recording of such gestures which are by nature dexterous and quick gestures, involving at the same time multimodal information (including bodily and hand movements, head movements, facial expressions and gaze). The team has acquired a Qualisys motion capture platform composed of eight Oqus400 cameras. We have defined several setups that allow us to capture different kinds of expressive gestures synchronized with video and sound. As hand movements and facial expressions present many occlusions and may give rise to noise and gaps in the data, it has been necessary to develop software tools to clean the data, correct it and fill the gaps of the markers' trajectories.

9.7 Corpora

Annotated Corpora

- PAROLE PUBLIQUE ("Free Speech") is a French corpus repository dedicated to pilot corpus studies for Human-Machine Dialogue, Augmentative and Alternative Communication (AAC) and Coreference Resolution.
- the ANCOR project deals with the study of all forms of anaphora and co-reference about the study of oral language. ANCOR_Centre, Accueil UBS and OTG corpora were built within the framework of this project.
- Various speech corpora including neutral speech with manual annotations (*Agnès* voice, FR, 7hours; *Elizabeth* voice, US, 7hours), expressive speech automatically annotated using audiobooks (10 hours), expressive speech with emotional content or even multi-speaker corpora such as BREF120.

10 Brief risk analysis

We take care on risks at three levels: scientific risks, management risks, and technical risks.

1. The main identified scientific risk is that each modality research axis keep on living as an independent activity. We will control this risk by means of common seminars and cross-domain research actions, to maintain synergies between the three modalities, and to favor projects or PhDs at the interface of these modalities. For example, a PhD has been started recently (end 2014) to incorporate facial expressions capabilities into talking heads and signing avatars; furthermore a joint bid is in preparation between text and speech modalities and will be submitted to the 2014 ANR call. We will take measures on a control of work progress; detailed and clear definition of our objectives and to focus on key issues, one at a time. A further issue is that we should be prepared to adapt the main team focus to alternative research routes if a direction turns out to be intractable or too far from the original objectives, given the capability of the team: staff size, time available for research activity (the staff is composed of only part time researchers), scientific know-how and development.
2. Managerial risks could include communication difficulties due to the geographical spread of the staff (Vannes, Lannion, Lorient). This risk could be partially overcome by extensive

use of visio-conferences or immersive rooms, with regular meetings. The lack of resources and/or staff allocated to the study of a given modality is also a difficulty that could affect the outcome in one of the three targeted application domains. We will ensure good level of communication to talk about rising problems and favor cooperation with national and international teams specialized in text, speech or gesture but also in cognitive and behavioral Sciences. Furthermore the funding being essentially ensured by contractualized sources, our main challenges could evolve in function of regional, national or international grants or fundings.

3. Technical risks may include difficulties in data acquisition or dedicated equipment management (cluster of machines, motion capture equipment, etc.). This risk is quite high, since the team does not rely on any permanent technical staff. Nevertheless, members of the team will potentially rely on centralized technical resources available in Rennes, e.g. for code forging purposes.

11 Appendix: team environment

We do not pretend providing here an exhaustive list of Laboratories or Teams around the world working in areas that are close to the topics put forward by EXPRESSION. We rather opted for a short list of focused teams, taking the risk to forgot some of them.

11.1 National environment

- IRCAM, UMR Sciences et Technologies de la Musique et du Son
  - Team sound analysis-synthesis
  - Expressive prosody
  - Real-Time Musical Interaction Team (IMTR)
  - `http://www.ircam.fr/54.html?&L=1`
- Laboratoire de Linguistique Formelle (LLF)
  - Work of Elisabeth Delais-Roussarie
  - Prosodic interfaces modelling, prosodic phonology and corpora
- Laboratoire de Phonétique et Phonologie (LPP)
  - Teams "TAL et phonétique" and "phonologie de corpus"
  - `http://lpp.in2p3.fr/article.php3?id_article=385`
- Laboratoire de Traitement et Communication de l'Information (LTCI), CNRS-TELECOM ParisTech, Greta Team
  - Embodied conversational agents; social and emotional behaviors in interaction
- Laboratoire d'Informatique de Grenoble (LIG)
  - Team GIPSA-lab (Grenoble), speech and cognition department, Gerard Bailly's Team
  - Team GETALP, work of Véronique Aubergé

  - Affective computing
  - http://www.liglab.fr/spip.php?article97

- Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI)
  - Group Image, Language, Space (AMI)
  - Group Cognition, Perception, and Usages (CPU)
  - http://www.limsi.fr/Scientifique/cpu/
  - Audio and Acoustics Group
  - Audio analysis and synthesis
  - http://groupeaa.limsi.fr/start

- Laboratoire d'Informatique de Nantes Atlantique (LINA)
  - Team TALN (Béatrice Daille)
  - Semantics of texts, opinion modelling, production of linguistic resources: corpora, lexicons, grammars
  - https://www.lina.univ-nantes.fr/?-TALN,68-.html

- Laboratoire d'Informatique de Paris-Nord (LIPN)
  - Team RCLN (Adeline Nazarenko)
  - Semantic Infoamrtion Retrieval, Pattern mining, Semantic Web
  - http://lipn.univ-paris13.fr/fr/rcln

- Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA)
  - Team PAROLE (Yves Laprie)
  - Acoustic-visual speech synthesis, multimodal speech processing, speech recognition, articulatory modeling
  - http://parole.loria.fr/index.php?lang=eng&fonction=accueil
  - Team ORPAILLEUR (Amedeo Napoli)
  - Text Mining, Formal Concept Analysis, Text-based ontology construction, Semantic Annotation and Semantic Web
  - http://orpailleur.loria.fr

- Voxygen
  - Expressive speech synthesis
  - http://voxygen.fr/

11.2 International environment

- Acapela
  - Expressive TTS, Multilingual, Emotive voices
  - http://www.acapela-group.com/
- AT&T Research, New Jersey, USA
  - AT&T Natural Voices^{TM} Text-to-Speech
  - http://www.research.att.com/projects/Natural_Voices/index.html?fbid=ZBlFNMevJXf
- Bristol Centre for Linguistics, England
  - production of linguistic resources, Parallel (translation) corpora,
  - http://www1.uwe.ac.uk/cahe/research/bristolcentreforlinguistics.aspx
- Centre for Speech Technology Research (CSTR), Edinburgh, UK
  - Speech synthesis, automatic speech recognition
  - http://www.cstr.ed.ac.uk
- Deutsche Telekom Laboratories, Germany
  - Speech perception, affective/emotional state
  - Felix Burkhardt: emotional Human Machine interaction, automatic speaker characteristics classification
  - Emotion profiles, modeling context and multimodality, emotional speech production, interaction modeling, natural language descriptions of emotions
  - http://www.laboratories.telekom.com/public/Deutsch/Publikationen/Pages/default.aspx and http://felix.syntheticspeech.de
- InfoMus Lab. Casa Paganini, Genova, Italy
  - Analysis of expressive gesture
  - http://www.infomus.org/index_eng.php
- Ivona (Amazon)
  - Multilingual TTS
  - http://www.ivona.com/en/
- Laboratory for Applied Ontology, Italy
  - conceptual modeling, knowledge representation, natural language processing, semantic web
  - http://www.loa.istc.cnr.it/
- Loquendo (Nuance)

  - Expressive TTS, Multilingual
  - http://www.nuance.fr/for-business/by-solution/customer-service-solutions/solutions-services/inbound-solutions/loquendo-small-business-bundle/text-to-speech/index.htm

- Carnegie Mellon University
  - Language Technologies Institute (LTI)
  - https://lti.cs.cmu.edu/

- Max Planck Institute, Group Multimedia Information Retrieval and Music Processing, Saarland University, Germany
  - Analysis of human motion, information retrieval
  - http://people.mpi-inf.mpg.de/~mmueller/

- MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), Cambridge, USA
  - Affective Computing Research Group
  - http://affect.media.mit.edu/
  - Humanoid Robotics group
  - Sociable machine project
  - http://www.ai.mit.edu/projects/sociable/overview.html

- Nagoya Institute of Technology, Japan
- Work of Keiichi Tokuda
  - HMM-based Text-To-Speech Synthesis (HTS)

- Ruth Aylett, University of Heriot Watt Edinburgh, UK
  - Affective computing, human / robot interaction
  - http://www.macs.hw.ac.uk/~ruth/bio.html

- Signal Analysis and Interpretation Lab, University of Southern California, Los Angeles, USA
  - Emotions Group
  - Emotion profiles, modeling context and multimodality, emotional speech production, interaction modeling, natural language descriptions of emotions
  - http://sail.usc.edu/emotion/index.php

- Sociable Agents Group (Center of Excellence Cognitive Interaction Technology – Bielefeld University – Faculty of Technology)
  - Interactive and adaptive virtual agents
  - http://www.techfak.uni-bielefeld.de/ags/soa/

- Technical University of Madrid, Department of Electronic Engineering, Spain
  - Speech Technology Group
  - Emotional speech synthesis
  - `http://lorien.die.upm.es`
- Toshiba Cambridge Research Laboratory, Cambridge, UK
  - Speech technology group
  - Rich, expressive synthesized speech, HTS, acoustic factorisation in speech synthesis with a system that allows the speaker and language to be controlled separately (a user can "speak" in several languages for personalized voice translation)
  - `http://www.toshiba.eu/eu/Cambridge-Research-Laboratory/Speech-Technology-Group/Speech-Technology-Group-Projects`
- TCTS Lab, Université de Mons, Belgique
  - Work of Thierry Dutoit on speech synthesis
  - `http://tcts.fpms.ac.be/~dutoit/`

Team publications (2009-present)

Books and book chapters

[ABB+12] J. Azé, N. Béchet, L. Berti-Équille, S. Guillaume, M. Roche, and F. Sais, *Mesurer et évaluer la qualité des données et des connaissances*. Éditions Hermann, 2012.

[BG09] A. Braffort and S. Gibet, *La Gestuelle*. Les Presses de l'Ecole des Mines, 2009, pp. 83–112.

[Gib10] S. Gibet, *Sensorimotor Control of Sound-Producing Gestures*, R. Godoy and M. L. Eds., Eds. Routledge publisher, 2010, pp. 212–237.

Phd theses

[Awa11] C. Awad, "Indexation et interrogation de base de données de mouvements pour l'animation d'humanoides virtuels", PhD thesis, 2011.

[Béc09b] N. Béchet, "Extraction et regroupement de descripteurs morpho-syntaxiques pour des processus de Fouille de Textes", PhD thesis, Université Montpellier 2, 2009.

[Bou09] A. Bouënard, "Synthesis of Music Performances: Virtual Character Animation as a Controller of Sound Synthesis", Anglais, PhD thesis, Dec. 2009.

[Dua12] K. Duarte, "Motion capture and avatars as portals for analyzing the linguistic structure of signed languages", PhD thesis, 2012.

[El 11] I. El Maarouf, "Formalisation de connaissances à partir de corpus : modélisation linguistique du contexte pour l'extraction automatique de relations sémantiques", Français, THESE, Dec. 2011.

[Ke14] G. Ke, "Mesures de comparabilité pour la construction assistée de corpus comparables bilingues thématiques", PhD thesis, 2014.

[Le 13a] S. Le Maguer, "Évaluation expérimentale d'un système statistique de synthèse de la parole, hts, pour la langue française", Français, Thesis, Université de Rennes 1, Jul. 2013.

[Le 13b] T. Le Naour, "Utilisation des relations spatiales pour l'analyse et l'édition de mouvement", PhD thesis, 2013.

[Lec10] G. Lecorvé, "Adaptation thématique non supervisée d'un système de reconnaissance automatique de la parole", PhD thesis, INSA de Rennes, 2010.

[Mes10] L. Mesbahi, "Transformation automatique de la parole, étude des transformations acoustiques", PhD thesis, Université de Rennes 1, 2010.

[Muh11] M. M. Muhammad Fuad, "Similarity Search in High-dimensional Spaces with Applications to Time Series Data Mining and Information Retrieval", Anglais, THESE, Feb. 2011.

Journal articles

[AVG12] J.-Y. Antoine, J. Villaneau, and J. Goulian, "Influence du genre applicatif sur la réalisation des extractions en dialogue oral : constantes et variations", *Langages*, no. 187, pp. 109–126, Sep. 2012.

[BCCC14] N. Béchet, P. Cellier, T. Charnois, and B. Crémilleux, "Fouille de motifs séquentiels pour la découverte de relations entre gènes et maladies rares", *Revue d'Intelligence Artificielle*, vol. 28, no. 2-3, pp. 245–270, 2014.

[BCPR10] N. Béchet, J. Chauché, V. Prince, and M. Roche, "CORPUS and WEB: Two Allies in Building and Automatically Expanding Conceptual Classes", *Informatica*, vol. 34, no. 3, 2010.

[BCPR14] ——, "How to combine text-mining methods to validate induced verb-object relations?", *Comput. Sci. Inf. Syst.*, vol. 11, no. 1, pp. 133–155, 2014.

[BRC09b] N. Béchet, M. Roche, and J. Chauché, "Comment valider automatiquement des relations syntaxiques induites", *Revue des Nouvelles Technologies de l'Information (RNTI) - numéro spécial @ EGC'2009*, 2009.

[BM12] N. Bonnel and P.-F. Marteau, "LNA: Fast Protein Classification Using A Laplacian Characterization of Tertiary Structure", *Transactions on Computational Biology and Bioinformatics, IEEE/ACM*, vol. 9 , Issue: 5, pp. 1451–1458, Oct. 2012. DOI: 10.1109/TCBB.2012.64.

[BGW12] A. Bouënard, S. Gibet, and M. M. Wanderley, "Hybrid Inverse Motion Control for Virtual Characters Interacting with Sound Synthesis - Application to Percussion Motion", *The Visual Computer Journal*, vol. 28, no. 4, pp. 357–370, Apr. 2012. DOI: 10.1007/s00371-011-0620-9.

[BWG10] A. Bouënard, M. M. Wanderley, and S. Gibet, "Gesture Control of Sound Synthesis: Analysis and Classification of Percussion Gestures", *Acta Acustica united with Acustica*, vol. 96, no. 4, pp. 668–677, 2010.

[BWGM11] A. Bouënard, M. M. Wanderley, S. Gibet, and F. Marandola, "Virtual Control and Synthesis of Music Performances: Qualitative Evaluation of Synthesized Timpani Exercises", *Computer Music Journal*, vol. 35, no. 3, pp. 57–72, Sep. 2011. DOI: 10.1162/COMJ_a_00069.

[CGM14] P. Carreno-Medrano, S. Gibet, and P. Marteau, "Synthèse de mouvements humains par des méthodes basées apprentissage : un état de l'art", *Revue Electronique Francophone d'Informatique Graphique*, vol. 8, no. 1, Jul. 2014.

[GCDL11a] S. Gibet, N. Courty, K. Duarte, and T. Le Naour, "The signcom system for data-driven animation of interactive virtual signers : methodology and evaluation", vol. 1, no. 1, p. 6, 2011.

[GGLS11] G. Gravier, C. Guinaudeau, G. Lecorvé, and P. Sébillot, "Exploiting speech for automatic tv delinearization: from streams to cross-media semantic navigation", *EURASIP Journal on Image and Video Processing*, vol. 2011, 2011.

[RL13] J. Ramos and C. Larboulette, "A muscle model for enhanced character skinning", *Journal of WSCG*, vol. 21, no. 2, pp. 107–116, Jul. 2013.

[KBR+12] R. Kessler, N. Béchet, M. Roche, J.-M. Torres-Moreno, and M. El-Bèze, "A hybrid approach to managing job offers and candidates", *Information Processing and Management*, vol. 48, no. 6, 2012.

[LBHR10] S. Laroum, N. Béchet, H. Hamza, and M. Roche, "Classification automatique de documents bruités à faible contenu textuel", *RNTI : Revue des Nouvelles Technologies de l'Information*, vol. E-18, no. Numéro spécial : Fouille de Données Complexes, 2010.

[LBHR11] ——, "Hybred: An OCR Document Representation for Classification Tasks", *IJCSI'2011: International Journal of Computer Science Issues*, vol. 8, no. Issue 3, No. 2, 2011.

[LCG12a] T. Le Naour, N. Courty, and S. Gibet, "Cinématique guidée par les distances", *Revue Électronique Francophone d'Informatique Graphique*, vol. Vol. 6, no. 1, pp. 15–25, 2012.

[LCG12c] ——, "Spatio-temporal coupling with the 3D+t motion Laplacian", *Revue Électronique Francophone d'Informatique Graphique*, vol. Vol. 6, no. 1, pp. 15–25, 2012.

[LBB10] D. Lolive, N. Barbot, and 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*, vol. 4, no. 3, pp. 571–581, 2010.

[Mar09] P.-F. Marteau, "Time Warp Edit Distance with Stiffness Adjustment for Time Series Matching", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 306–318, Feb. 2009. DOI: 10.1109/TPAMI.2008.76.

[MBM12] P.-F. Marteau, N. Bonnel, and G. Ménier, "Discrete Elastic Inner Vector Spaces with Application in Time Series and Sequence Mining", *IEEE Transactions on Knowledge and Data Engineering*, pages, Jun. 2012. DOI: 10.1109/TKDE.2012.131.

[MG14b] P.-F. Marteau and S. Gibet, "On Recursive Edit Distance Kernels with Application to Time Series Classification", *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, Jun. 2014, in publishing process. DOI: 10.1109/TNNLS.2014.2333876.

Conferences

[ABB+09] P. Alain, N. Barbot, V. Barreaud, L. Blin, O. Boëffard, L. Charonnat, A. Choumane, A. Delhay, S. L. Maguer, D. Lolive, T. Moudenc, and G. Vidal, "A multi-agent platform for multimodal pervasive applications", in *Proceedings of the Conference on the Networked and Electronic Media (Nem)*, Saint Malo, France, 2009.

[AVL14a] J.-Y. Antoine, J. Villaneau, and A. Lefeuvre, "Weighted krippendorff's alpha is a more reliable metrics for multi-coders ordinal annotations: experimental studies on emotion, opinion and coreference annotation", in *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, Gothenburg, Sweden: Association for Computational Linguistics, Apr. 2014, pp. 550–559.

[AGVL09] J.-Y. Antoine, J. Goulian, J. Villaneau, and M. Le Tallec, "Word Order Phenomena in Spoken French : a Study on Four Corpora of Task-Oriented Dialogue and its Consequences on Language Processing", in *Proc. 5th Corpus Linguistics Conference*, Liverpool, Royaume-Uni, Jun. 2009, actes electroniques.

[ALV11] J.-Y. Antoine, M. Le Tallec, and J. Villaneau, "Evaluation de la détection des émotions, des opinions ou des sentiments : dictatute de la majorité ou respect de la diversité d'opinions ?", in *Actes de la conférence TALN'2011*, vol. 2, Montpellier, France, Jun. 2011, 6 p.

[AdLG10] M. Aubry, P. de Loor, and S. Gibet, "Enhancing robustness to extrapolate synergies learned from motion capture", in *Proc. of the International Conference on Computer Animation and Social Agents (CASA)*, Saint Malo, 2010.

[ACL+14] M. Avanzi, G. Christodoulides, D. Lolive, E. Delais-Roussarie, and N. Barbot, "Towards the adaptation of prosodic models for expressive text-to-speech synthesis", in *Proceedings of the Interspeech conference*, 2014.

[ACD+09a] C. Awad, N. Courty, K. Duarte, T. Le Naour, and S. Gibet, "A Combined Semantic and Motion Capture Database for Real-Time Sign Language Synthesis", in *Proc. of the 9th Int. Conference on Intelligent Virtual Agent (IVA 2009)*, Amsterdam, Pays-Bas, 2009, pp. 432–438.

[ACG09] C. Awad, N. Courty, and S. Gibet, "A Database Architecture For Real-Time Motion Retrieval", in *Proc. of the 7th International Workshop on Content-Based Multimedia Indexing (CBMI 2009)*, I. CS, Ed., Chania, Greece, France, 2009, pp. 225–230.

[BBB+11a] N. Barbot, V. Barreaud, O. Boëffard, L. Charonnat, A. Delhay, S. Le Maguer, and D. Lolive, "Towards a Versatile Multi-Layered Description of Speech Corpora Using Algebraic Relations", in *Proceedings of Interspeech*, 2011, pp. 1501–1504.

[BBD12] N. Barbot, O. Boëffard, and A. Delhay, "Comparing performance of different set-covering strategies for linguistic content optimization in speech corpora", in *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*, 2012.

[BMP13a] N. Barbot, L. Miclet, and H. Prade, "Analogical proportions and the factorization of information in distributive lattices ", in *Proceedings of the Conference on Concept Lattices and Their Applications*, 2013.

[BMP13b] ——, "Proportions analogiques et factorisation de l'information dans les treillis distributifs", in *Journées d'Intelligence Artificielle Fondamentale. GDR I3*, 2013, à paraître.

[Béc09a] N. Béchet, "Description d'un protocole d'évaluation automatique comme alternative à l'évaluation humaine. Application à la validation de relations syntaxiques induites", in *Evaluation des Méthodes d'Extraction de Connaissances dans les Données, Atelier de EGC'09*, 2009.

[BAL12] N. Béchet, M.-A. Aufaure, and Y. Lechevallier, "Construction et peuplement de structures hiérarchiques de concepts dans le domaine du e-tourisme", in *Actes de IC2011*, 2012.

[BCCC12b] N. Béchet, P. Cellier, T. Charnois, and B. Crémilleux, "Fouille de motifs séquentiels pour la découverte de relations entre gènes et maladies rares", in *Acte des 23es Journées Francophones d'Ingénierie des Connaissances - IC 2012*, ser. IC 2012, 2012.

[BCC+13] N. Béchet, P. Cellier, T. Charnois, B. Crémilleux, and S. Quiniou, "SDMC : un outil en ligne d'extraction de motifs séquentiels pour la fouille de textes", in *Actes de la Conférence Francophone sur l'Extraction et la Gestion des Connaissances (EGC'13)*, 2013.

[BC12] N. Béchet and M. Csernel, "Comparing Sanskrit Texts for Critical Editions: the sequences move problem", in *13th Internationlal Conference on Intelligent Text Processing and Computational Linguistics*, 2012.

[BR10] N. Béchet and M. Roche, "How to Expand Dictionaries with Web-Mining Techniques", in *COGALEX'10: Cognitive Aspects of the Lexicon*, 2010.

[BRC09a] N. Béchet, M. Roche, and J. Chauché, "A Hybrid Approach to Validate Induced Syntactic Relations", in *The 2009 IEEE International Symposium on Mining and Web, in conjunctiion with IEEE AINA'09*, 2009.

[BRC09c] ——, "Corpus et Web : deux alliés pour la construction de l'enrichissement automatique de classes conceptuelles", in *Toth'09 : Terminologie & Ontologie : Théories et Applications*, 2009.

[BRC09d] ——, "Towards the Selection of Induced Syntactic Relations", in *ECIR'09: European Conference on Information Retrieval*, ser. LNCS, Springer-Verlag, 2009.

[BCLL12] O. Boëffard, L. Charonnat, S. Le Maguer, and 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.

[BCL+12] O. Boëffard, L. Charonnat, S. Le Maguer, D. Lolive, and G. Vidal, "Vers une annotation automatique de corpus audio pour la synthèse de parole", in *Actes de la conférence conjointe JEP-TALN-RECITAL*, 2012, pp. 731–738.

[BMM09b] N. Bonnel, G. Ménier, and P.-F. Marteau, "Proactive Uniform Data Replication by Density Estimation in Apollonian P2P Networks", in *Proceedings of Second International Conference on Data Management in Grid and Peer-to-Peer Systems (GLOBE)*, S. B. /. Heidelberg, Ed., ser. Lecture Notes in Computer Science, vol. 5697, Autriche: Springer Berlin / Heidelberg, Sep. 2009, pp. 60–71. DOI: 10.1007/978-3-642-03715-3_6.

[BGW09] A. Bouënard, S. Gibet, and M. M. Wanderley, "Hybrid Motion Control combining Inverse Kinematics and Inverse Dynamics Controllers for Simulating Percussion Gestures", in *Proc. of the International Conference on Computer Animation and Social Agents (CASA)*, Amsterdam, The Netherlands, 2009, pp. 17–20.

[BWG09a] A. Bouënard, M. M. Wanderley, and S. Gibet, "Advantages and Limitations of Simulating Percussion Gestures for Sound Synthesis", in *Proc. of the International Computer Music Conference (ICMC)*, 2009, pp. 255–261.

[BWG09b] ——, "Analysis of Timpani Preparatory Gesture Parameterization", in *Proc. of the International Gesture Workshop*, 2009, pp. 61–62.

[BBB+11b] S. Bringay, N. Béchet, F. Bouillot, P. Poncelet, M. Roche, and M. Teisseire, "Towards an On-Line Analysis of Tweets Processing", in *Database and Expert Systems Applications*, A. Hameurlain, S. W. Liddle, K.-D. Schewe, and X. Zhou, Eds., ser. LNCS, vol. 6861, Springer, 2011.

[CGLM14] P. Carreno-Medrano, S. Gibet, C. Larboulette, and P.-F. Marteau, "A combined semantic and motion capture database for real-time sign language synthesis", in *IVA*, 2014.

[CLL14a] J. Chevelu, G. Lecorvé, and 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.

[CLL14b] ——, "ROOTS : un outil pour manipuler facilement, efficacement et avec cohérence des corpus annot'es de séquences", in *Actes des 30es Journées d'Étude sur la Parole (JEP)*, 2014.

[CG10] N. Courty and S. Gibet, "Why is the creation of a virtual signer challenging computer animation?", in *Motion In Games 2010*, 2010, pp. 290–300.

[DLY+14] E. Delais-Roussarie, D. Lolive, H. Yoo, N. Barbot, and O. Rosec, "Adapting prosodic chunking algorithm and synthesis system to specific style: the case of dictation", in *Proceedings of the Interspeech conference*, 2014.

[DG10a] K. Duarte and S. Gibet, "Corpus design for signing avatars", in *Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, Valletta, Malta: European Language Resources Association (ELRA), 2010.

[DG10c] ——, "Heterogeneous data sources for signed language analysis and synthesis: the signcom project", in *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta: European Language Resources Association (ELRA), 2010.

[DG10d] ——, "Reading between the signs: how are transitions built in signed languages?", in *Theretical Issues in Sign Language Research (TILSR 2010), Indiana, USA*, 2010.

[EV12a] I. El Maarouf and J. Villaneau, "A French Fairy Tale Corpus syntactically and semantically annotated.", in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, N. C. (Chair), K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk, and S. Piperidis, Eds., Istanbul, Turquie, May 2012, pages.

[EV12b] ——, "Parenthetical Classification for Information Extraction", in *Proceedings of COLING 2012: Posters*, T. C. 2. O. Committee, Ed., Mumbai, Inde, Dec. 2012, pp. 297–308.

[EVSD09] I. El Maarouf, J. Villaneau, F. Saïd, and D. Duhaut, "Comparing Child and Adult Language : Exploring Semantic constraints", in *ICMI-MLMI'09 Workshop on Child, Computer and Interaction*, Cambridge, MA, États-Unis, 2009, pages.

[GDC11] S. Gibet, K. Duarte, and N. Courty, "Signing Avatars: Linguistic and Computer Animation Challenges", in *First International Symposium on Sign Language Translation and Avatar Technology (SLTAT) 2011*, Berlin, Germany, Jan. 2011.

[GM09] S. Gibet and P.-F. Marteau, "Approximation of Curvature and Velocity for Gesture Segmentation and Synthesis", in *Proceedings of International Gesture Workshop*, ser. Lecture Notes in Computer Science/ Computer Science, vol. ISBN: 978-3-540-92864-5, Portugal: Springer, Dec. 2009, pp. 13–23. DOI: 10.1007/978-3-540-92865-2_2.

[GL14a] D. Guennec and D. Lolive, "Unit Selection Cost Function Exploration Using an A* based Text-to-Speech system", in *proceedings of the TSD conference*, 2014.

[GL14b] ——, "Utilisation d'un algorithme A* pour l'analyse de la sélection d'unité en synthèse de la parole", in *JEP - 30ème édition des Journées d'Etudes sur la Parole*, Le Mans, France, Jun. 2014.

[HGB13] L. Hamon, S. Gibet, and S. Boustila, "Édition interactive d'énoncés en langue des signes française dédiée aux avatars signeurs", in *TALN - 20ème conférence du Traitement Automatique du Langage Naturel 2013*, Sables d'Olonne, France, Jun. 2013.

[IBC+12] D. Imseng, H. Bourlard, H. Caesar, P. Garner, G. Lecorvé, and A. Nanchen, "Mediaparl: bilingual mixed language accented speech database", in *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*, 2012, pp. 263–268.

[ATL13] F. J. Alcon Palazon, D. Travieso, and C. Larboulette, "Influence of dynamic wrinkles on the perceived realism of real-time character animation", in *Proceedings of the International Conference on Computer Graphics, Visualization and Computer Vision (WSCG)*, M. M.Oliveira and V. Skala, Eds., ser. Annual Conference Series, Full Papers, Eurographics, c/o University of West Bohemia, Czech Republic: Vaclav Skala - Union Agency, Jun. 2013, pp. 95–103.

[KMG12] J.-F. Kamp, G. Ménier, and S. Gibet, "Une interface gestuelle pour l'apprentissage de la rythmique", in *Actes de la conférence RFIA 2012*, Lyon, France, Jan. 2012, pages.

[KM14] G. Ke and P.-F. Marteau, "Co-clustering of bilingual datasets as a mean for assisting the construction of thematic bilingual comparable corpora", in *The 9th edition of the Language Resources and Evaluation Conference, LREC 2014*, Reykjavik, Islande, May 2014, pp.

[KMM14] G. Ke, P.-F. Marteau, and G. Ménier, "Variations on quantitative comparability measures and their evaluations on synthetic French-English comparable corpora", in *The 9th edition of the Language Resources and Evaluation Conference, LREC 2014*, Reykjavik, Islande, May 2014, pp.

[KBT+09a] R. Kessler, N. Béchet, J.-M. Torres-Moreno, M. Roche, and M. El-Bèze, "Job Offer Management: How Improve the Ranking of Candidates", in *ISMIS'09: International Symposium on Methodologies for Intelligent Systems*, 2009.

[KBT+09b] ——, "Profilage de candidatures assisté par relevance Feedback", in *TALN'09 : Traitement Automatique des Langues Naturelles*, 2009.

[LQD13] C. Larboulette, P. Quesada Barriuso, and O. Dumas, "Burning paper: simulation at the fiber's level", in *Proceedings of the ACM SIGGRAPH Conference on Motion in Games*, ser. ISBN: 978-1-4503-2546-2, http://dx.doi.org/10.1145/2522628.2522906, ACM New York, NY, USA ©2013, ACM Press, Nov. 2013, pp. 25–30.

[LBB12] S. Le Maguer, N. Barbot, and O. Boëffard, "Evaluation segmentale du système de synthèse hts pour le français", in *Actes de la conférence conjointe JEP-TALN-RECITAL*, 2012, pp. 569–576.

[LDB+14a] S. Le Maguer, E. Delais-Roussarie, N. Barbot, M. Avanzi, O. Rosec, and D. Lolive, "Algorithme de découpage en groupes prosodiques pour la dictée par l'usage de synthèse vocale", in *JEP - 30ème édition des Journées d'Etudes sur la Parole*, Le Mans, France, Jun. 2014.

[LDB+14b] ——, "Prosodic chunking algorithm for dictation with the use of speech synthesis", in *Proc. of Speech Prosody*, Dublin, Ireland, 2014.

[LCG12b] T. Le Naour, N. Courty, and S. Gibet, "Fast Motion retrieval with the distance input space", in *Motion in Games - 5th International Conference, MIG 2012,*, M. Kallmann and K. E. Bekris, Eds., vol. 7660, Rennes, France, 2012, pp. 362–365.

[LAVD11a] M. Le Tallec, J.-Y. Antoine, J. Villaneau, and D. Duhaut, "Affective Interaction with a Companion Robot for Hospitalized Children: a Linguistically based Model for Emotion Detection", in *Proceedings of the 5th Language and Technology Conference (LTC'2011)*, Poznan, Pologne, Nov. 2011, 6 pages.

[LAV+10] M. Le Tallec, J.-Y. Antoine, J. Villaneau, A. Savary, and A. Syssau, "Détection hors contexte des émotions à partir du contenu linguistique d'énoncés oraux : le système EmoLogus", in *TALN2010, Montreal : Canada(2010)*, Montreal, Canada, 2010, p. 167.

[LSJ+10] M. Le Tallec, S. Saint-Aimé, C. Jost, J. Villaneau, J.-Y. Antoine, S. Letellier-Zarshenas, B. Le Pévédic, and D. Duhaut, "From speech to emotional interaction: EmotiRob project", in *3rd International Conference on Human-Robot Personal Relationships*, Leiden, Pays-Bas: Springer, 2010, p. 8.

[LVA+09] M. Le Tallec, J. Villaneau, J.-Y. Antoine, A. Savary, and A. Syssau-Vaccarella, "Détection des émotions à partir du contenu linguistique d'énoncés oraux : application à un robot compagnon pour enfants fragilisés", in *Actes TALN 2009*, Senlis, France, Jun. 2009, p. 90.

[LVA+10] M. Le Tallec, J. Villaneau, J.-Y. Antoine, A. Savary, and A. Syssau-Vaccarella, "Emologus - A Compositional Model of Emotion Detection based on the Propositionnal Content of Spoken Utterances", in *Proceedings Text Speech and Dialog 2010*, ser. LNCS/LNAI, vol. 6231, Brno, Tchèque, République: Springer, 2010, 8 pages.

[LDHM12a] G. Lecorvé, J. Dines, T. Hain, and P. Motlicek, "Impact du degré de supervision sur l'adaptation à un domaine d'un modèle de langage à partir du web", in *Actes de la conférence conjointe JEP-TALN-RECITAL*, 2012, pp. 193–200.

[LDHM12b] ——, "Supervised and unsupervised web-based language model domain adaptation", in *Proceedings of Interspeech*, 2012, pp. 131–134.

[LGS09] G. Lecorvé, G. Gravier, and P. Sébillot, "Constraint selection for topic-based MDI adaptation of language models", in *Proceedings of Interspeech*, 2009, pp. 368–371.

[LGS10] ——, "L'adaptation thématique d'un modèle de langue fait-elle apparaître des mots thématiques ?", in *Actes des 28è Journées d'Étude sur la Parole (JEP)*, 2010.

[LGS11] ——, "Automatically finding semantically consistent n-grams to add new words in LVCSR systems", in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 4676–4679.

[LM12] G. Lecorvé and P. Motlicek, “Conversion of recurrent neural network language models to weighted finite state transducers for automatic speech recognition”, in *Proceedings of Interspeech*, 2012, pp. 131–134.

[LBB09] D. Lolive, N. Barbot, and O. Boëffard, “An evaluation methodology for prosody transformation systems based on chirp signals”, in *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, Brighton, UK, 2009, pp. 2635–2638.

[MG14a] P.-F. Marteau and S. Gibet, “Down-Sampling coupled to Elastic Kernel Machines for Efficient Recognition of Isolated Gestures”, in *ICPR 2014, 22nd International Conference on Pattern Recognition. August 24-28, 2014, Waterfront, Stockholm, Sweden.*, IAPR, Ed., Stockholm, Suède: IEEE, Aug. 2014, pp.

[MM13] P.-F. Marteau and G. Ménier, “Similarités induites par mesure de comparabilité : signification et utilité pour le clustering et l’alignement de textes comparables”, in *Actes 20e conférence sur le Traitement Automatique des Langues Naturelles (TALN’2013)*, Les Sables d’Olonne, France, Jun. 2013, pp. 515–522.

[MM09a] G. Ménier and P.-F. Marteau, “Tabu Split and Merge for the Simplification of Polygonal Curves”, in *SMC’09: Proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics*, IEEE, Ed., ser. IEEE SMC, vol. isbn 978-1-4244-2793-2, États-Unis: IEEE, Oct. 2009, pp. 1322–1327. DOI: 10.1109/ICSMC.2009.5346240.

[MBJ12a] L. Miclet, N. Barbot, and B. Jeudy, “A lattice of sets of alignments built on the common subwords in a finite language”, in *Proceedings of the International Conference on Grammatical Inference*, 2012, pp. 164–176.

[MBJ12b] ——, “Analogical proportions in a lattice of sets of alignments built on the common subwords in a finite language”, Proceedings of the 1st ECAI Workshop on Similarity and Analogy-based Methods in AI, 2012.

[MPG11] L. Miclet, H. Prade, and D. Guennec, “Looking for Analogical Proportions in a Formal Concept Analysis Setting”, in *Proceedings of the Conference on Concept Lattices and Their Applications*, 2011, pp. 295–307.

[MM09b] A. Mouton and P.-F. Marteau, “Exploiting routing information encoded into backlinks to improve topical crawling”, in *Proceedings of the International Conference of Soft Computing and Pattern Recognition, SOCPAR 2009*, IEEE, Ed., ser. IEEE, vol. SocPAR 2009, Malaisie: IEEE, Dec. 2009, pp. 659–664. DOI: 10.1109/SoCPaR.2009.129.

[MM10a] M. M. Muhammad Fuad and P.-F. Marteau, “Enhancing the Symbolic Aggregate Approximation Method Using Updated Lookup Tables”, in *Knowledge-Based and Intelligent Information and Engineering Systems,*, ser. Lecture Note in Computer Sciences, vol. Volume 6276/2010, Royaume-Uni: Springer Verlag, Sep. 2010, pp. 420–431. DOI: 10.1007/978-3-642-15387-7_46.

[MM10b] ——, “Fast Retrieval of Time Series Using a Multi-resolution Filter with Multiple Reduced Spaces”, in *Advanced Data Mining and Applications*, ser. Lecture Notes in Computer Science, Chine, Nov. 2010, pp. 137–148. DOI: 10.1007/978-3-642-17316-5_13.

[MM10c] ——, "Multi-resolution Approach to Time Series Retrieval", in *IDEAS'2010*, ser. ACM International Conference Proceeding Series, France: ACM, Aug. 2010, pp. 136–142. DOI: 10.1145/1866480.1866501.

[MM10d] ——, "Speeding-up the Similarity Search in Time Series Databases by Coupling Dimensionality Reduction Techniques with a Fast-and-dirty Filter", in *Proceedings of the Fourth IEEE International Conference on Semantic Computing (ICSC2010)*, États-Unis: IEEE CS, Sep. 2010, pp. 101–104. DOI: 10.1109/ICSC.2010.34.

[MM10e] ——, "Symbolic Aggregate Approximation Method Using Updated Lookup Tables", in *Proceeding of the 14th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems – KES 2010*, vol. Lecture Notes in Computer Sciences, Royaume-Uni, Sep. 2010, pp. 1–12.

[MM10f] ——, "Towards a Faster Symbolic Aggregate Approximation Method", in *Proceeding of the 5th International Conference on Software and Data Technologies*, Grèce, Jul. 2010, pp. 1–6.

[MLS+14a] J. Muzerelle, A. Lefeuvre, E. Schang, J.-Y. Antoine, A. Pelletier, D. Maurel, I. Eshkol, and J. Villaneau, "Ancor_centre, a large free spoken french coreference corpus: description of the resource and reliability measures", in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, N. C. (Chair), K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds., Reykjavik, Iceland: European Language Resources Association (ELRA), May 2014, ISBN: 978-2-9517408-8-4.

[VA09] J. Villaneau and J.-Y. Antoine, "Deeper spoken language understanding for man-machine dialogue on broader application domains: a logical alternative to concept spotting.", in *Proceedings of EACL 2009 WORKSHOP ON SEMANTIC REPRESENTATION OF SPOKEN LANGUAGE*, Athènes, Grèce, Mar. 2009, pp. 50–57.

[YLD+14] H. Yoo, S. Le Maguer, E. Delais-Roussarie, N. Barbot, and D. Lolive, "Evaluation d'un algorithme de chunking appliqué à la dictée", in *JEP - 30ème édition des Journées d'Etudes sur la Parole*, Le Mans, France, Jun. 2014.

Technical reports

[PRMM11a] Y. Péron, F. Raimbault, G. Ménier, and P.-F. Marteau, "On the detection of inconsistencies in rdf data sets and their correction at ontological level", Tech. Rep., Jun. 2011.

Other

[AJG09] M. Aubry, F. Julliard, and S. Gibet, "Modeling joint synergies to synthesize realistic movements", in *Gesture in Human-Computer Interaction and Simulation, Lecture Notes in Artificial Intelligence, LNAI*, Springer Verlag, 2009, pp. 231–242.

[BMM09a] N. Bonnel, P.-F. Marteau, and G. Ménier, "Parallel Random Apollonian Networks", Apr. 2009.

[BHB+13] F. Bouillot, P. N. Hai, N. Béchet, S. Bringay, D. Ienco, S. Matwin, P. Poncelet, M. Roche, and M. Teisseire, "How to Extract Relevant Knowledge from Tweets?", in *Communications in Computer and Information Science*, 2013.

[BGH+12] R. Brun, S. Gibet, L. Hamon, F. Lefebvre-Albaret, and A. Turki, *The sign3d project*, http://http://sign3d.websourd.org/, 2012.

[GDL+10] S. Gibet, K. Duarte, T. Le Naour, N. Courty, F. Multon, S. Donikian, P. Dalle, and J. Foucher, *The signcom project*, http://www-valoria.univ-ubs.fr/signcom/en/, 2010.

[GMD12a] S. Gibet, P.-F. Marteau, and K. Duarte, "Toward a Motor Theory of Sign Language Perception", in *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication, LNCS*, E. Efthimiou, G. Kouroupetroglou, and S.-E. Fotinea, Eds., Springer Berlin Heidelberg, 2012, Vol. 7206, 161–172, ISBN: 978-3-642-34181-6. DOI: 10.1007/978-3-642-34182-3_15.

[GBC+13] M. Grandgeorge, P. Blanchet, Y. Chevalier, D. Duhaut, M. Hausberger, A. Lemasson, B. Le Pévédic, F. Poirier, J. Peeters, F. Pugniere - saavedra, M. Rinn, and J. Villaneau, "Modélisation interdisciplinaire de l'intercompréhension dans les interactions", in *Interactions et Intercompréhension : une approche comparative Homme-Homme, Animal-Homme-Machine et Homme-Machine*, ser. echanges, EME editions, Apr. 2013, pp. 103–123, ISBN: ISBN 978-2-8066-0859-8.

[HG09] A. Héloir and S. Gibet, "A qualitative and quantitative characterisation of style in sign language gestures", in *Gesture in Human-Computer Interaction and Simulation, GW 2007, Lecture Notes in Artificial Intelligence, LNAI*, Lisboa, Portugal: Springer Verlag, 2009, pp. 122–133.

[KM13] G. Ke and P.-F. Marteau, "Improving the clustering or categorization of bi-lingual data by means of comparability mapping", Oct. 2013.

[LGT+13] F. Lefebvre-Albaret, S. Gibet, A. Turki, L. Hamon, and R. Brun, *Overview of the Sign3D Project High-fidelity 3D recording, indexing and editing of French Sign Language content*, Chicago, États-Unis, Oct. 2013.

[Mar11] P.-F. Marteau, "Discrete Time Elastic Vector Spaces", Jan. 2011.

[MBJ14] L. Miclet, N. Barbot, and B. Jeudy, "Analogical Proportions in a Lattice of Sets of Alignments Built on the Common Subwords in a Finite Language", in *Computational Approaches to Analogical Reasoning: Current Trends, Studies in Computational Intelligence*, H. Prade and G. Richard, Eds., Springer-Verlag Berlin Heidelberg, 2014, pp. 245–260. DOI: 10.1007/978-3-642-54516-0_10.

[PRMM11b] Y. Péron, F. Raimbault, G. Ménier, and P.-F. Marteau, "Supervised tracking and correction of inconsistencies in rdf data", May 2011.

References

[Abe95] M. Abe, "Speaking Styles: Statistical Analysis and Synthesis by a Text-to-Speech System", in *Progress in Speech Synthesis*, J. P. Van Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg, Eds., 1995, ch. 39, pp. 495–510.

[ABB+08] P. Alain, N. Barbot, O. Boëffard, J. Chevelu, and A. Delhay, "Evaluation de méthodes de réduction de corpus linguistiques", in *Actes des XXVIIèmes Journées d'Etudes sur la Parole (JEP)*, 2008.

[AB14] H. Alatrista Salas and N. Béchet, "Fouille de texte : une approche séquentielle pour découvrir des relations spatiales", in *Proceedings of CerGEO'2014 (Construction, l'enrichissement et l'exploitation de ressources GEOgraphiques pour l'analyse de données), EGC'2014 workshop*, RNTI, 2014, pp. 1–8.

[All76] J. Allen, "Synthesis of speech from unrestricted text", *Proceedings of the IEEE*, vol. 64, no. 4, pp. 433–442, 1976.

[AFO03] O. Arikan, D. A. Forsyth, and J. F. O'Brien, "Motion synthesis from annotations", *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 402–08, Jul. 2003.

[AI06] O. Arikan and L. Ikemoto, *Computational Studies of Human Motion: Tracking and Motion Synthesis*. Now Publishers Inc, 2006.

[AC14] A. Aristidou and Y. Chrysanthou, "Feature extraction for human motion indexing of acted dance performances", in *GRAPP 2014 - Proceedings of the 9th International Conference on Computer Graphics Theory and Applications, Lisbon, Portugal, 5-8 January, 2014.*, 2014, pp. 277–287. DOI: 10.5220/0004662502770287.

[AP08] R. Artstein and M. Poesio, "Inter-coder agreement for computational linguistics", *COMPUTATIONAL LINGUISTICS*, vol. 34, no. 4, pp. 555–596, 2008.

[AVAR06] N. Audibert, D. Vincent, V. Aubergé, and O. Rosec, "Expressive speech synthesis: evaluation of a voice quality centered coder on the different acoustic dimensions", in *Proceedings of Speech Prosody*, vol. 2006, 2006, pp. 525–528.

[ACD+09] C. Awad, N. Courty, K. Duarte, T. L. Naour, and S. Gibet, "A combined semantic and motion capture database for real-time sign language synthesis", in *IVA*, 2009, pp. 432–438.

[Béc01] F. Béchet, "Lia_phon : un système complet de phonétisation de textes", *Traitement Automatique des Langues (TAL)*, vol. 42, no. 1, pp. 47–67, 2001.

[BCCC12] N. Béchet, P. Cellier, T. Charnois, and B. Crémilleux, "Discovering linguistic patterns using sequence mining", in *CICLing (1)*, 2012, pp. 154–165.

[BCC+12] N. Béchet, P. Cellier, T. Charnois, B. Cremilleux, and M.-C. Jaulent, "Sequential pattern mining to discover relations between genes and rare diseases", in *Proceedings of Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, IEEE, 2012, pp. 1–6.

[BSHR05] G. Beller, D. Schwarz, T. Hueber, and X. Rodet, "A hybrid concatenative synthesis system on the intersection of music and speech", in *Proceedings of Journées d'Informatique Musicale*, 2005, pp. 41–45.

[BDVJ03] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model", *Journal of Machine Learning Research*, vol. 3, no. 2, pp. 1137–1155, Feb. 2003.

[BDHS11] G. Berio, A. D. Leva, M. Harzallah, and J. M. Sacco, "Competence management over social networks through dynamic taxonomies.", Anglais, in *Encyclopedia of KM2.0 : Organizational Models and Enterprise Strategies*, IGI Global, 2011, pp. 103–120.

[BR09] D. Bernhardt and P. Robinson, "Detecting emotions from connected action sequences", in *Proceedings of the the International Visual Informatics Conference*, 2009.

[BN08] M. Bisani and H. Ney, "Joint-sequence models for grapheme-to-phoneme conversion", *Speech Communication*, 2008.

[Bla07] A. W. Black, "Speech Synthesis for Educational Technology", in *SLATE*, 2007, pp. 78–81.

[BT94] A. W. Black and P. Taylor, "CHATR: a generic speech synthesis system", in *Proceedings of the 15th conference on Computational linguistics*, 1994, pp. 983–986.

[BRBd02] C. Blouin, O. Rosec, P. C. Bagshaw, and C. d'Alessandro, "Concatenation cost calculation and optimisation for unit selection in tts", ser. IEEE Workshop on Speech Synthesis, 2002, pp. 231–234, ISBN: 0780373952.

[Bon04] F. Bonchi, "On closed constrained frequent pattern mining", in *In Proceedings IEEE Int. Conf. on Data Mining ICDM'04*, Press, 2004, pp. 35–42.

[BH00] M. Brand and A. Hertzmann, "Style machines", in *ACM SIGGRAPH 2000*, 2000, pp. 183–192.

[Bre92] A. Breen, "Speech synthesis models: a review", *Electronics & communication engineering journal*, vol. 4, no. 1, pp. 19–31, 1992.

[BNS02] M. Bulut, S. S. Narayanan, and A. K. Syrdal, "Expressive speech synthesis using a concatenative synthesizer", in *Proceedings of ICSLP*, 2002, pp. 1265–1268.

[CBBD07] J. Chevelu, N. Barbot, O. Boéffard, and A. Delhay, "Lagrangian relaxation for optimal corpus design", in *Proceedings of ISCA Speech Synthesis Workshop*, 2007, pp. 211–216.

[CBBD08] J. Chevelu, N. Barbot, O. Boëffard, and A. Delhay, "Comparing set-covering strategies for optimal corpus design", in *Proceedings of LREC*, 2008.

[CCZB00] D. CHI, M. Costa, L. Zhao, and N. Badler, "The emote model for effort and shape", in *SIGGRAPH'00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., 2000, pp. 173–182.

[CL06] Y.-R. Chien and J.-S. Liu, "Learning the stylistic similarity between human motions", in *Proceedings of the Second international conference on Advances in Visual Computing - Volume Part I*, ser. ISVC'06, 2006, pp. 170–179.

[CHB13] R. Chulyadyo, M. Harzallah, and G. Berio, "Core Ontology based Approach for Treating the Flatness of Automatically Built Ontology", Anglais, in *Proceedings of KEOD*, Portugal, 2013, 316:323.

[CRK07] R. A. J. Clark, K. Richmond, and S. King, "Multisyn: open-domain unit selection for the festival speech synthesis system", *Speech Communication*, vol. 49, no. 4, pp. 317–330, 2007.

[CBSP00] A. Conkie, M. C. Beutnagel, A. K. Syrdal, and E. Philip, "Preselection of candidate units in a unit selection-based text-to-speech synthesis system", ser. In ICSLP, vol. 3, 2000, pp. 314–317.

[CG07] E. Crane and M. Gross, *Motion Capture and Emotion: Affect Detection in Whole Body Movement*, ser. Affective Computing and Intelligent Interaction, ACII, Lecture Notes in Computer Science. Springer Verlag, 2007, vol. 4738, pp. 95–101, In Proceedings of ACII,

[dGel06] B. de Gelder, "Toward a biological theory of emotional body language", *Biological Theory*, vol. 1, pp. 130–132, 2006.

[DGWS06] G. Demenko, S. Grocholewski, A. Wagner, and M. Szymanski, "Prosody annotation for corpus based speech synthesis", in *Proceedings of the Eleventh Australasian International Conference on Speech Science and Technology*, 2006, pp. 460–465.

[Don98] E. M. Donovan R. E.and Eide, "The ibm trainable speech synthesis system", *International Conference on Spoken Language Processing*, 1998.

[DG10] K. Duarte and S. Gibet, "Corpus design for signing avatars", in *In Proceedings of the 4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, Valletta, Malta: European Language Resources Association (ELRA), 2010.

[Dut97] T. Dutoit, "High-quality text-to-speech synthesis: An overview", *Journal of Electrical and Electronics Engineering*, vol. 17, no. 1, pp. 25–36, 1997.

[EAB+04] E. Eide, A. Aaron, R. Bakis, W. Hamza, M. Picheny, and J. Pitrelli, "A corpus-based approach to expressive speech synthesis", in *Fifth ISCA Workshop on Speech Synthesis*, 2004.

[EL13] A. Elgammal and C.-S. Lee, "Homeomorphic manifold analysis (hma): generalized separation of style and content on manifolds", *Image and Vision Computing*, vol. 31, no. 4, pp. 291–310, 2013.

[Eri05] D. Erickson, "Expressive speech: production, perception and application to speech synthesis", *Acoustical Science and Technology*, vol. 26, no. 4, pp. 317–325, 2005.

[Gal09] P. Galaher, "Individual differences in nonverbal behavior: dimensions of style", in *Journal of Personality and Social Psychology*, vol. 51, 2009, pp. 133–145.

[Gan13] A. Gangemi, "A comparison of knowledge extraction tools for the semantic web", in *ESWC*, 2013, pp. 351–366.

[GR94] C. Gerard and C. Rigaut, "Patterns prosodiques et intentions des locuteurs : le rôle crucial des variables temporelles dans la parole", *Le Journal de Physique IV*, vol. 04, no. C5, pp. 505–508, May 1994, ISSN: 1155-4339. DOI: 10.1051/jp4:19945107.

[GBHK12] T. Gherasim, G. Berio, M. Harzallah, and P. Kuntz, “Problems impacting the quality of automatically built ontologies”, Anglais, in *Proceedings of KESE*, vol. 949, France: Ceur, 2012.

[GHBK13] T. Gherasim, M. Harzallah, G. Berio, and P. Kuntz, “A comparative analysis of some approaches for automatic construction of ontologies from textual resources”, Anglais, in *Advances In Knowledge Discovery and Management*, Springer, 2013, pp. 177–201.

[GCDL11] S. Gibet, N. Courty, K. Duarte, and T. Le Naour, “The signcom system for data-driven animation of interactive virtual signers : methodology and evaluation”, vol. 1, no. 1, 2011.

[GMD12] S. Gibet, P.-F. Marteau, and K. Duarte, “Toward a motor theory of sign language perception”, *Human-Computer Interaction and Embodied Communication, GW 2011*, vol. 7206, pp. 161–172, 2012.

[GP13] D. Govind and S. R. M. Prasanna, “Expressive speech synthesis: a review”, *International Journal of Speech Technology*, pp. 1–24, 2013.

[GMHP04] K. Grochow, S. L. Martin, A. Hertzmann, and Z. Popović, “Style-based inverse kinematics”, *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 522–531, 2004.

[GCF10] M. Gross, E. Crane, and B. Fredrickson, “Methodology for assessing bodily expression of emotion”, in *Journal of Nonverbal Behaviour*, vol. 34, 2010, pp. 223–248.

[HMBP05] B. Hartmann, M. Mancini, S. Buisine, and C. Pelachaud, “Design and evaluation of expressive gesture synthesis for embodied conversational agents”, in *AAMAS*, 2005, pp. 1095–1096.

[HBO12] M. Harzallah, G. Berio, and A. L. Opdahl, “New perspectives in ontological analysis: guidelines and rules for incorporating modelling languages into ueml”, *Inf. Syst.*, vol. 37, no. 5, pp. 484–507, 2012.

[HCGM06a] A. Héloir, N. Courty, S. Gibet, and F. Multon, “Temporal alignment of communicative gesture sequences”, *Computer Animation and Virtual Worlds*, vol. 17, pp. 347–357, 2006.

[HCGM06b] A. Heloir, N. Courty, S. Gibet, and F. Multon, “Temporal alignment of communicative gesture sequences”, *Journal of Visualization and Computer Animation*, vol. 17, no. 3-4, pp. 347–357, 2006.

[Her03] A. Hertzmann, “Machine learning for computer graphics: a manifesto and tutorial”, in *Computer Graphics and Applications, 2003. Proceedings. 11th Pacific Conference on*, IEEE, 2003, pp. 22–36.

[HPP05] E. Hsu, K. Pulli, and J. Popović, “Style translation for human motion”, in *ACM Transactions on Graphics (TOG)*, ACM, vol. 24, 2005, pp. 1082–1089.

[HAA+96] X. Huang, A. Acero, J. Adcock, H.-W. Hon, J. Goldsmith, J. Liu, and M. Plumpe, “Whistler: a trainable text-to-speech system”, in *International Conference on Spoken Language Processing*, 1996, pp. 2387–2390.

[IFJ11] I. Illina, D. Fohr, and D. Jouvet, “Multiple pronunciation generation using grapheme-to-phoneme conversion based on conditional random fields”, in *Proceedings of the International Conference on Speech and Computer (SPECOM)*, 2011.

[IMK+04] Y. Irie, S. Matsubara, N. Kawaguchi, Y. Yamaguchi, and Y. Inagaki, "Speech Intention Understanding based on Decision Tree Learning", in *Interspeech*, 2004, pp. 2185–2188.

[IAML04] I. Iriondo, F. Alias, J. Melenchon, and M. A. Llorca, "Modeling and Synthesizing Emotional Speech for Catalan Text-to-Speech Synthesis", in *Affective Dialogue Systems*, 2004, pp. 197–208.

[ISA07] I. Iriondo, J. C. Socoró, and F. Alías, "Prosody modelling of spanish for expressive speech synthesis", in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, IEEE, vol. 4, 2007, pp. IV–821.

[Jel76] F. Jelinek, "Continuous speech recognition by statistical methods", *Proc. of the IEEE*, vol. 64, no. 4, pp. 532–556, Apr. 1976.

[JS02] M. Jilka and A. K. Syrdal, "The AT&T german text-to-speech system: realistic linguistic description", in *International Conference on Spoken Language Processing*, 2002.

[KTN+04] H. Kawai, T. Toda, J. Ni, M. Tsuzaki, and K. Tokuda, "Ximera: a new tts from atr based on corpus-based technologies", in *ISCA ITRW on Speech Synthesis*, 2004, pp. 179–184.

[Ken80] A. Kendon, "Gesticulation and speech two aspects of the process of utterance", in *The Relation Between Verbal and Nonverbal Communication*, 1980, pp. 207–227.

[Kop10] S. Kopp, "Social resonance and embodied coordination in facetoface conversation with artificial interlocutors", *Speech Communication*, vol. 52, no. 6, pp. 587–597, 2010.

[KW04] S. Kopp and I. Wachsmuth, "Synthesizing multimodal utterances for conversational agents", *Journal of Visualization and Computer Animation*, vol. 15, no. 1, pp. 39–52, 2004.

[Kri04] K. Krippendorff, "Reliability in content analysis: some common misconceptions and recommendations.", *Human Communication Research*, vol. 30, no. 3, pp. 411–433, 2004.

[LM11] A. Lacheret-Dujour and M. Morel, "Modéliser la prosodie pour la synthèse à partir du texte : Perspectives sémantico-pragmatiques", in *Au commencement était le verbe. Syntaxe, sémantique et cognition*, note 23, F. Neveu, P. Blumenthal, and N. Le Querler, Eds., Peter Lang, 2011, pp. 299–325.

[LFV+11] P. Lanchantin, S. Farner, C. Veaux, G. Degottex, N. Obin, G. Beller, F. Villavicencio, T. Hueber, D. Schwartz, S. Huber, *et al.*, "Vivos voco: a survey of recent research on voice transformations at ircam", in *International Conference on Digital Audio Effects (DAFx)*, 2011, pp. 277–285.

[LAVD11] M. Le Tallec, J.-Y. Antoine, J. Villaneau, and D. Duhaut, "Affective Interaction with a Companion Robot for Hospitalized Children: a Linguistically based Model for Emotion Detection", Anglais, in *Proceedings of the 5th Language and Technology Conference (LTC'2011)*, Poznan, Pologne, 2011, 6 pages.

[LSS+06] Y. Liu, E. Shriberg, A. Stolcke, D. Hillard, M. Ostendorf, and M. Harper, "Enriching speech recognition with automatic detection of sentence boundaries and disfluencies", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1526–1540, 2006.

[Mal87] V. Maletik, *Body, Space, Expression : The Development of Rudolf Laban?s Movement and Dance Concepts*. Walter de Gruyter Inc., 1987.

[MBC+06] J. B. Marino, R. E. Banchs, J. M. Crego, A. de Gispert, P. Lambert, J. A. Fonollosa, and M. R. Costa-Jussà, "N-gram-based machine translation", *Computational Linguistics*, vol. 32, no. 4, pp. 527–549, 2006.

[MTKI96] T. Masuko, K. Tokuda, T. Kobayashi, and S. Imai, "Speech synthesis using HMMs with dynamic features", in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, IEEE, 1996, pp. 389–392.

[MHM12] T. Matsubara, S.-H. Hyon, and J. Morimoto, "Real-time stylistic prediction for whole-body human motions", *Neural Networks*, vol. 25, pp. 191–199, 2012.

[McN92] D. McNeill, *Hand and Mind - What Gestures Reveal about Thought*. Chicago, IL: The University of Chicago Press, 1992.

[Mik12] T. Mikolov, "Statistical language models based on neural networks", PhD thesis, Brno University of Technology, 2012.

[MDK+11] T. Mikolov, A. Deoras, S. Kombrink, L. Burget, and J. Cernocky, "Empirical evaluation and combination of advanced language modeling techniques", in *Proceedings of the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH 2011)*, 2011, pp. 605–608.

[MKB+10] T. Mikolov, M. Karafiat, L. Burget, J. Cernocky, and S. Khudanpur, "Recurrent neural network based language model", in *Proc. of the Conf. of the Intl Speech Communication Association (Interspeech)*, 2010, pp. 1045–1048.

[MBS09] M. Müller, A. Baak, and H.-P. Seidel, "Efficient and robust annotation of motion capture data", in *Proceedings of the ACM SIGGRAPH Eurographics Symposium on Computer Animation*, New Orleans, LA, Aug. 2009, pp. 17–26.

[MA96] I. Murray and J. Arnott, "Synthesizing emotions in speech: is it time to get excited?", in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, vol. 3, Ieee, 1996, pp. 1816–1819, ISBN: 0-7803-3555-4. DOI: 10.1109/ICSLP.1996.607983.

[NLHP98] R. Ng, L. Lakshmanan, J. Han, and A. Pang, "Exploratory mining and pruning optimizations of constrained associations rules", in *Proceedings of SIGMOD'98*, 1998, pp. 13–24.

[OBHM12] A. L. Opdahl, G. Berio, M. Harzallah, and R. Matulevicius, "An ontology for enterprise and information systems modelling", *Applied Ontology*, vol. 7, no. 1, pp. 49–92, 2012.

[PT09] W. Pan and L. Torresani, "Unsupervised hierarchical modeling of locomotion styles", in *Proceedings of the 26th Annual International Conference on Machine Learning*, ACM, 2009, pp. 785–792.

[PHW02] J. Pei, J. Han, and W. Wang, "Mining sequential patterns with constraints in large databases", ACM Press, 2002, pp. 18–25.

[PHW07] ——, "Constraint-based sequential pattern mining: the pattern-growth methods", *Journal of Intelligent Information Systems*, vol. 28, pp. 133–160, 2 2007, ISSN: 0925-9902.

[PP10] T. Pejsa and I. S. Pandzic, "State of the art in example-based motion synthesis for virtual characters in interactive applications", in *Computer Graphics Forum*, Wiley Online Library, vol. 29, 2010, pp. 202–226.

[Pel09] C. Pelachaud, *Studies on Gesture Expressivity for a Virtual Agent*, 1. 2009, vol. 63, pp. 630–639.

[PBE+06] J. F. Pitrelli, R. Bakis, E. M. Eide, R. Fernandez, W. Hamza, and M. A. Picheny, "The ibm expressive text-to-speech synthesis system for american english", *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1099–1108, Jul. 2006, ISSN: 1558-7916. DOI: 10.1109/TASL.2006.876123.

[QDMS01] S. Quazza, L. Donetti, L. Moisa, and P. L. Salza, "Actor: a multilingual unit-selection speech synthesis system", in *ISCA ITRW on Speech Synthesis*, 2001, p. 209.

[QBD04] C. Quirk, C. Brockett, and W. B. Dolan, "Monolingual machine translation for paraphrase generation.", in *EMNLP*, 2004, pp. 142–149.

[RJ93] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*. Prentice hall Englewood Cliffs, New Jersey, 1993.

[RSHM09] A. R. F. Rebordao, M. a. M. Shaikh, K. Hirose, and N. Minematsu, "How to Improve TTS Systems for Emotional Expressivity", in *Interspeech*, 2009, pp. 524–527.

[RBC98] C. Rose, B. Bodenheimer, and M. F. Cohen, "Verbs and adverbs: multidimensional motion interpolation using radial basis functions", *IEEE Computer Graphics and Applications*, vol. 18, pp. 32–40, 1998.

[Sch09] M. Schröder, "Expressive speech synthesis: past, present, and possible futures", in *Affective information processing*, Springer, 2009, pp. 111–126.

[Sch01] M. Schröder, "Emotional Speech Synthesis : A Review", in *Proceedings of Eurospeech*, 2001.

[SMLR05] B. Schuller, R. Müller, M. Lang, and G. Rigoll, "Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles", in *Proceedings of Interspeech*, 2005, pp. 805–808.

[SG02] H. Schwenk and J.-L. Gauvain, "Connectionist language modeling for large vocabulary continuous speech recognition", in *Proc. of IEEE Intl Conf on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, 2002, pp. 765–768.

[Shr94] E. Shriberg, "Preliminaries to a theory of speech disfluencies", PhD thesis, University of California, Berkeley, California, USA, 1994.

[SC99] F. Song and W. B. Croft, "A general language model for information retrieval", in *Proceedings of the eighth international conference on Information and knowledge management*, ACM, 1999, pp. 316–321.

[SS96] A. Stolcke and E. Shriberg, "Statistical language modeling for speech disfluencies", in *Proceedings of the IEEE Intl Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, Atlanta, Georgia, USA, May 1996, pp. 405–408.

[SFK+05] A. Stolcke, L. Ferrer, S. Kajarekar, E. Shriberg, and A. Venkataraman, "Mllr transforms as features in speaker recognition", in *in Proceedings of the 9th European Conference on Speech Communication and Technology*, Citeseer, 2005.

[SCK06] V. Strom, R. Clark, and S. King, "Expressive Prosody for Unit-selection Speech Synthesis", in *Proceedings of the International Conference on Speech Communication and Technology (Interspeech)*, 2006.

[SNH03] D. Sundermann, H. Ney, and H. Hoge, "Vtln-based cross-language voice conversion", in *Automatic Speech Recognition and Understanding, 2003. ASRU'03. 2003 IEEE Workshop on*, IEEE, 2003, pp. 676–681.

[TYMK07] N. Takashi, J. Yamagishi, T. Masuko, and T. Kobayashi, "A style control technique for HMM-based expressive speech synthesis", *IEICE TRANSACTIONS on Information and Systems*, vol. 90, no. 9, pp. 1406–1413, 2007.

[Tay09] P. Taylor, *Text-to-speech synthesis*. Cambridge UK: Cambridge University Press, 2009, vol. 1, ISBN: 9780511515057.

[TBC98a] P. Taylor, A. Black, and R. Caley, "The architecture of the Festival speech synthesis system", in *The Third ESCA Workshop in Speech Synthesis*, Citeseer, 1998, pp. 147–151.

[TBC98b] *The architecture of the Festival speech synthesis system*, ser. The Third ESCA Workshop in Speech Synthesis, Citeseer, 1998, pp. 147–151.

[05] "The Gaston Tool for Frequent Subgraph Mining", *Electronic Notes in Theoretical Computer Science*, vol. 127, no. 1, pp. 77–87, 2005.

[TZ02] K. Tokuda and H. Zen, "An HMM-based speech synthesis system applied to English", in *Speech Synthesis, 2002.*, 2002.

[THB06] L. Torresani, P. Hackney, and C. Bregler, "Learning motion style synthesis from perceptual observations", in *Advances in Neural Information Processing Systems*, 2006, pp. 1393–1400.

[VB11] N. Vincze and Y. Bestgen, "An automatic procedure for extending lexical norms by means of the analysis of word co-occurrences in texts", *TAL*, vol. 52, no. 3, pp. 191–216, 2011.

[Vox13] Voxygen, *Voxygen Online Speech Synthesis System, http://voxygen.fr*, 2013.

[Wal98] H. Wallbott, "Bodily expression of emotion", in *European Journal of Social Psychology*, vol. 28, 1998, pp. 879–896.

[WFH07] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Multifactor gaussian process models for style-content separation", in *Proceedings of the 24th international conference on Machine learning*, ACM, 2007, pp. 975–982.

[WH04] J. Wang and J. Han, "Bide: efficient mining of frequent closed sequences", in *Proceedings of the 20th International Conference on Data Engineering*, ser. ICDE '04, Washington, DC, USA: IEEE Computer Society, 2004, pp. 79–, ISBN: 0-7695-2065-0.

[WHLW06] C.-H. Wu, C.-C. Hsia, T.-H. Liu, and J.-F. Wang, "Voice conversion using duration-embedded bi-hmms for expressive speech synthesis", *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 4, pp. 1109–1116, 2006.

[YH02] X. Yan and J. Han, "Gspan: graph-based substructure pattern mining", in *Proceedings of the 2002 IEEE International Conference on Data Mining*, ser. ICDM '02, Washington, DC, USA: IEEE Computer Society, 2002, pp. 721–, ISBN: 0-7695-1754-4.

[YHA03] X. Yan, J. Han, and R. Afshar, "Clospan: mining closed sequential patterns in large databases", in *SDM*, 2003.

[ZTB09] H. Zen, K. Tokuda, and A. W. Black, "Statistical parametric speech synthesis", *Speech Communication*, vol. 51, no. 11, pp. 1039–1064, 2009, ISSN: 0167-6393.