HAL
archives-ouvertes.fr

SASI: perspectives for a socio-affectively intelligent HRI dialog system

Yuko Sasa, Véronique Aubergé

▶ To cite this version:

Yuko Sasa, Véronique Aubergé. SASI: perspectives for a socio-affectively intelligent HRI dialog system. 1st Workshop on "Behavior, Emotion and Representation: Building Blocks of Interaction", Oct 2017, Bielefeld, Germany. <hal-01615470>

HAL Id: hal-01615470
https://hal.inria.fr/hal-01615470

Submitted on 12 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SASI: perspectives for a socio-affectively intelligent HRI dialog system

Yuko Sasa
LIG, CNRS UMR 5217, UGA
Grenoble, France
Yuko.Sasa@univ-grenoble-alpes.fr

Véronique Aubergé
LIG, CNRS UMR 5217, UGA
Grenoble, France
Veronique.Auberge@univ-grenoble-alpes.fr

ABSTRACT
Situated face-to-face interactions are dynamic processes, building our social role and our relationships. In the present study, the bond resulting from these communicational processes is called the socio-affective glue, which is expected to be the element to recognise in our ASR prototype. This system is motivated to follow the relation changing instead of being accurate to distinguish the lexical cues and their semantics, to adapt the possible feedbacks giving the impression of a socio-affective intelligence in the perspective of a global incremental and iterative dialogue system (SASI) that can ensure the communication continuity. This approach uses the observations of the EEE corpus, composed by spontaneous and ecological interactions between socio-isolated elderly and a Smart Home's butler robot, which is activating automation through vocal commands.

Author Keywords
Spontaneous and ecological corpus, socio-affective glue, human-robot interaction, Smart Home, ASR, incremental dialog system.

ACM Classification Keywords
Algorithms, Experimentation, Human Factors, Languages, Measurement.

INTRODUCTION
The robot is a machine artefact which focuses on the dynamic relationship between the human and the machine, how and which parameters, and the usefulness of the relation itself, inscribing the HRI strongly in ethical debates. The crucial challenge to build the relationship is developed "attaching" dialogue in understanding the benefits and the risks of such human-machine attachment. Since the 70's, the spoken dialogue systems had two approaches: 1) conversational vision with a "human metaphor"; and 2) command-based systems with an "interface metaphor" [1]. The first one is focused on the modelling of what people say, inducing a small vocabulary and grammar while the second proposes the possible variations of "saying something" in a restricted interactional context but with a particular motivation to obtain a universal speech interface [2]. Most of them are following the dialogue acts model [3][4] as introduced with DAMSL [5]. Nevertheless, it seems the "dialogue" has focused on the "adjacency pairs" approach defined as in [6] with forward-looking and backward-looking exchanges. They consequently concentrated on the accuracy to response to the lexical object or the socially codified contents' semantics. In traditional dialogue systems, this kind of information seems to follow the processing in a pipeline with temporal constraints inducing a delay of the system's response. The spoken dialogue systems are globally following successively: an audio input, an Automatic Speech Recognition (ASR), a Natural Language Understanding (NLU), a Dialog Management (DM), a Natural Language Generation (NLG) and a Text-To-Speech synthesis (TTS) [7][8]. Nevertheless, the development of incremental dialogue systems increased the possibilities of asynchronous processing of each component, as seen in the Eve agents [9], NUMBERS systems [10] or the Pentomino of the Inpro project [11]. The incremental methods are mainly adopting two approaches. On the one hand, the methods use handcrafted, supervised or reinforcement learning [12][13], which need a costing collection of input data preparation or simulated data [14][15]. On the other hand, the methods test the results in a real-life condition, possibly using the produced data on live [16]. As [10] mention, a global and continuous "self-monitoring" analyses and adjusts our communicational behaviours during the interlocutor's hearing, as well as our own talking time. This idea also motivates the developers of incremental dialogue systems, as "overtly or covertly self-repairs" can also be observed. As the lexical and interaction semantics stay ambiguous in natural language and even for humans, another approach might have a more relational and global approach to communication, making use of the advancement of an incremental dimension.

The present paper proposes a possible approach with this perspective, basing firstly on the observation from the collected EmOz Elderly Expressions – EEE – corpus. The data gather natural and ecological interactions between a Smart Home butler robot and some socio-isolated elderly giving automation vocal commands. In previous works, [17] have shown that spontaneous nonverbal sounds can be strongly significant in human interaction, and [18] make emerge perceptively the progressive attaching values of a selection of these sounds, partially dependent on the culture [19]. The EEE corpus is collected, induced by a strong hypothesis named after the "socio-affective glue" defined as the relational consequence of the dynamic

communicational processes. It is supposed to be gradually built and modified by using specific prosodic features based on vocal micro-expressions, which represent the robot's vocal feedbacks manipulated with the – EmOz – wizard of Oz protocol. Secondly, we propose to change the paradigm of classical ASR approach by trying to recognise the relation instead of words/phrases. This modified perspective is motivated to underlie a whole incremental and iterative dialogue system, - SASI – Socio-Affective Speech Interaction, by trying to change the incremental properties of the communication.

THE SOCIO-AFFECTIVE GLUE BUILDING

In face-to-face interaction, the dialogue dynamicity is a challenging issue to understand how we continuously adapt and try to keep in touch with our interlocutors. A channel is so to be built, a one that can carry the communicational tools we use to interact. These objects are supposed to modify continually this channel shapes, as well as our social role, both influencing each other. In the speech, the non-lexical features exist inside and outside the turn-takings. These markers seem to rhythm and to ensure the dialogue continuity. These short vocal features known as grunts, mind markers, affect bursts, fillers, interjections [20][21], nonverbal cues [22][23] or paralinguistic tools [24] were showed to build the same kind of information as the socio-affective prosody on sentences [25][26]. Thus, we supposed these vocal micro-expressions to be "pure prosody events" carrying the same communicative functions. In the present study, these vocal micro-expressions, are expected to be able to build an interactional channel between two interlocutors needing as a first condition, a social role. The dynamic communication process is so supposed to perpetually changing the speakers' relational statement named after the "socio-affective glue" [27]. This relationship building process is proposed to depend on mechanisms like the ones observed in grooming, which would rely on an altruistic dimension, entirely independent of the dominance dimension. For this "glueing" process [27] observation, we selected vocal events (but mostly their prosody) from the spontaneous E-Wiz corpus of French micro-expressions [28]. They were multimodally labelled [18][29], perceptually validated in intracultural and intercultural conditions [30][19] to assess the informational value and communicational functions [31]. These sounds built within a particular socio-affective "pure" prosody are supposed to have a "glueing" effect, following an order depending on the degree of their production and communicative intentionality control. In this study corpus, these vocal events are the tools to be used in a recreated micro-ecological context as the only communication tool of a robot to control their production and graduation [31]. This robot was interacting with socio-isolated elderly for whom this bond building process is damaged, a contrastive situation, which eases the observation of this socio-affective glue dynamics.

DATA: EEE CORPUS

Corpus collection experimental settings

The complete setting description is available in [34]. The experiment follows a methodology developed in a living-lab environment co-constructed with industrial partners (Awabot robotics company) and social partners (caregiving services and institution) to induce spontaneous and ecological situated interactions. A sophisticated and controlled scenario tricks the elderly. It leads them progressively on a one-on-one interaction with the robot. The EmOz Wizard of Oz platform controls the robot and the automation entirely. The sounds emitted by the robot are feedbacks to the subject's vocal commands activating the Smart Home automation. They are supposed to induce a glueing effect, changing the relational statement, and so the way the subjects interact with the robot. The data is transcribed, aligned, labelled and auto-annotated (i.e. auto-annotation is an externally controlled way to label socio-affective information of the interactional utterances).

EEE corpus characteristics

The –EEE– Elderly EmOz Expressions corpus collected 24 French subjects interactions (4 men and 23 women, in the same proportion as healthy elderly in French institutions). A panel of 20 subjects aged between 68 to 93 years old is entirely exploitable for now. The social isolation was on criteria to select the subjects, by consulting the caregivers, social partners or kinfolks. This isolation is an important methodological parameter to observe the glueing process. All the elderly follow the French frailty scale, on a GIR5 and GIR6 degree, which is corresponding to fragile, helped but relatively independent persons without cognitive impairments. Table 1 illustrates the quantitative characteristics of the actual EEE corpus.

Table 1. The EEE corpus quantitative description.

The EEE corpus data composition	Quantity
Total labeled elderly's speech utterances	11253
Elderly's spontaneous commands	1321
Vocal micro-expressions	1888
Not proposed commands-like utterances	263
Direct non-command utterances to emox	3820
Human-human utterances including emox	3327
Collected HRI phase speech (with silences)	14:38:25
Total duration of experiments inside Domus	46:22:42

Key observations for the dialog systems perspectives

In this corpus, the robot's micro-expressions effects were observed through the subject's vocal micro-events, following their vocal commands utterances [33]. The key points of this data set are:

- the proposed voice commands are paraphrased through the experiment, while the instructions ask to follow the exact command form;
- the prosody of subjects and the robot converges. Singular points in echolalia and imitations appeared punctually. Moreover, the elderly voice quality changes with an increasing breathiness and pitch through the experiment;
- the aged people's vocal micro-expressions quantity increase and their temporal organisation change;
- global similarities characterise the vocal features dynamics and other multimodal cues (proxemics, posture, gestures, moves and facial expressions);
- all the previous changes seem to follow in some way the socio-affective glue statement, verified on the qualitative auto-annotations labels given by the elderly themselves;

Finally, the subjects' tolerance differs towards the induced robots "error", regarding a certain "degree" and statement of glue. They showed a tendency to reject the robot if the error occurred in a lower glue level (by skipping, abandoning the failing command or expressing a negative opinion on the tool). On the other hand, if this error occurred on a higher glue statement the elderly recovered the robot's mistake and strengthen the glue.

FOCUS ON A DIFFERENT PARADIGM FOR ASR UNDERLYING THE SASI DIALOG SYSTEM

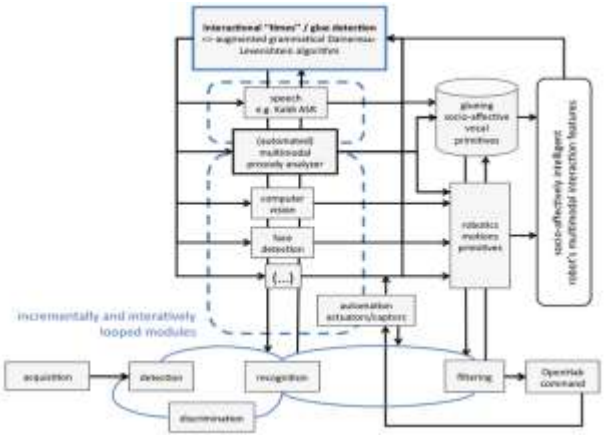

Figure 1. Perspective scheme of SASI dialog system.

An ASR prototype recognizing the glue statement

Complex architectures build dialogue systems where the ASR is a restricted module integrated into this whole system. However, if the objects recognised by the ASR are different, it might transform the entire paradigm of the dialogue. Our study proposes to explore this possibility, placing the ASR on the central position of this global system. The Kaldi toolkit [34] was to prototype an ASR on DNNs trained on the BREF120 corpus, to build an online decoding system. Its input parameters are raw (non-adapted nor normalised) log-Mel filter bank coefficients on 40-dimensional features, calculated from a 7-frames context window, with additional 600 dimensions i-vector characterising the speaker. The neural network is trained on five hidden layers with a 400-dimensional p-norm input, 4000 for the output and associated to 12000 sub-classes. An offline version was also developed to test the system on the EEE corpus dataset. In this version, the acoustic model is based on a GMM previously trained with the BREF120 corpus, using triphones, and with an adaptation based on LDA, MLLT, SAT and fMLLR methods.

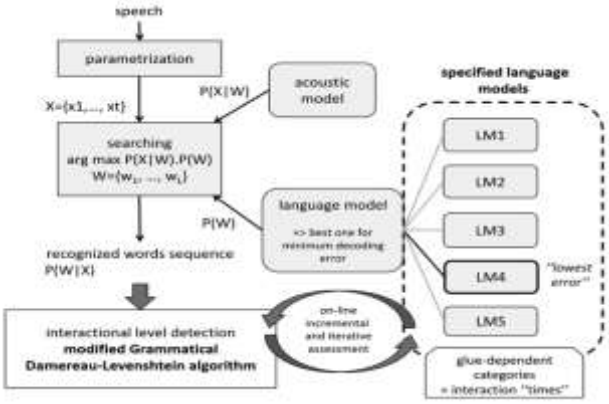

Figure 2. Socio-affective glue-dependent ASR scheme.

The Figure 1 left part shows a classical training/adaptation method, common to every ASR. It exists various and discussed techniques, but the difference on which we would like to focus on is the aim of the language model (LM). In general, this LM handles the decoding as accurately as possible the exact message. In the present work, we would like to recognise some interactional "times". The frontiers would be defined by the moments the elderly "switches" into a different socio-affective glue statement. However, the subjects' vocal commands change morpho-syntactically, showing some regular patterns associated to "relational timings". Instead of having the feature description in a single LM, the Figure 2 proposes to develop multiple LM, while each of them illustrates one glue instance. In fact, some speech features are supposed to be peculiar to an LM. In a unique LM, the possible words combinations increase as well as the ambiguity to decode. The choice of one LM instead of another could be:

- deduced by its decoding efficiency
- a predictor of the relational statement shared with the interlocutor.
- the proposed ASR prototype, integrated into a perspective of a dialogue system presents two components:
- the LMs deduced from our EEE corpus observations
- adaptation of a grammatical Damereau-Levenshtein algorithm based on the predicted ASR output command.

Glue-dependent language models

The ASR prototype uses 5 LM illustrating different "glue levels", g1 to g5 represented in Figure 3. The levels choices

are arbitrary from the analysis of the morpho-syntactical changes and the associated auto-annotations. We split the dataset into two sets, k1 and k2 for a cross learning-test. Firstly, learning was done with k1 then test on k2, then inversely. A gall including all the data was set to interpolate the kx,g1 to kx,g5 specific models with the kx,gall model. Both models kx,gy and kx,gall associate an interpolation with a model learned on the OpenSubtitles corpus, to handle a larger vocabulary. We gave a weight of 90% to the specialised kx,gy and kx,gall models, and 10% on the OpenSubtitles based model. Every model uses trigrams learned with the SRILM toolkit.

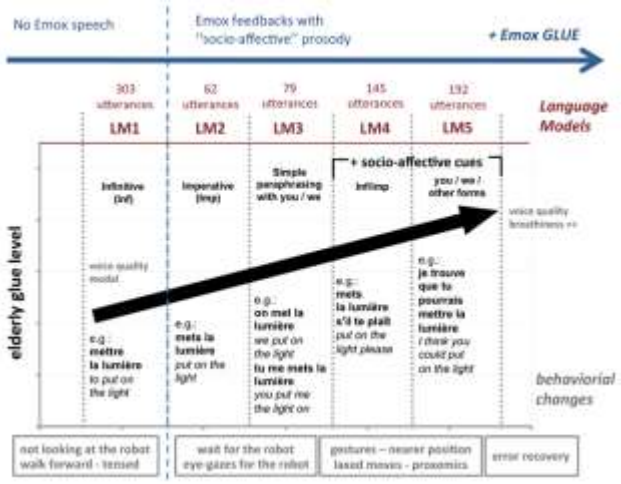

Figure 3. Socio-affective glue-dependent ASR scheme.

Currently, the model selection uses a perplexity measure defining how a model predicts a sample. While the perplexity is small, the sample is close to the model. Meanwhile, this perplexity can support comparisons between samples, but it is not possible to compare them between models. This models comparison define which LM is the closest one to the input sample by training a classifier on the boosting algorithm, the AdaBoost using the Bonzaiboost tool. The algorithm uses the following parameters measured with the SRILM tool: word (number of word in the utterance), logprob (the probability log), ppl (normalised perplexity counting all the tokens) and ppl1 (normalised perplexity without the statement's ending tags). The F-Measure applied for the different models showed the best score while all of the four classifiers were used, and in without interpolation. The WER (Word Error Rate) calculation proposes a cross-validation and a "tricked"-validation of the k1, k2 samples (k1 training- k1 test, k2 training- k2 test). Because of the limited size of the training data extract from the EEE corpus, the results for the cross-validation won't be relevant. However, we focus some primary results on the "tricked"-validation, to test if the even idea of having multiple specialised models can be relevant. Table 2 shows the interpolated data result with the Open-Subtitles corpus.

Table 2. WER results for the "tricked"-validation on interpolated specialized LMs and a global LM.

Test \ LMs	g_1_OpenSu	g_2_OpenSu	g_3_OpenSu	g_4_OpenSu	g_5_OpenSu	g_{all}_OpenSu
$k_x g_1$	6,70%	**50,19%**	**30,10%**	**14,40%**	**23,25%**	9,39%
$k_x g_2$	**56,25%**	25,57%	**57,95%**	**44,32%**	**56,25%**	35,23%
$k_x g_3$	**44,16%**	**48,72%**	13,68%	**37,89%**	**25,64%**	16,81%
$k_x g_4$	**44,61%**	**60,56%**	**57,54%**	22,84%	**50,22%**	25,75%
$k_x g_5$	**56,22%**	**60,21%**	**52,22%**	**51,62%**	18,31%	19,37%
					15,56%	**18,11%**

Instead of the dataset's size, the specialised LMs show globally a lower WER: 1) while the LM corresponds to gx with the same "glue level rank" compared to the other LM (in green); on the best-specialized LM compared to the global model based on gall (in blue). This result is not incompatible with our approach, and it is even encouraging. On the cross-validation test the first LM1, based on the greater number of utterances (same "infinitive form" as the proposed command form) showed the same tendency. However, the efficiency for the other specialised LMs was not that clear. On the one hand, the results depend on the dataset's size, but on the contrary, improvements have been made for each specialised LM themselves (even create others). In fact, the EEE corpus new parts pointed out different patterns and graduation in the existing one. Moreover, this first attempt is not sufficiently accurate on the "socio-affective cues", appearing very frequently for the LM4 and LM5. Currently, they are not distinct, while they concern various vocal micro-expressions with different value and function. So revising the models on these factors might be a possible perspective for future work.

A modified Grammatical Damerau-Levenshtein algorithm to match the socio-affective glue level

In parallel, the dialogue system is expected to have a module incrementally calculating a "glue statement", where the first input will be the ASR's decoding results. The modified Damerau-Levenshtein algorithm stands for the glue level calculation. The classical algorithm is an edit distance calculator between two compared utterances. In our conjecture, the first one is the model (the "imposed commands form"); the second one is the tested production (the "spontaneous elderly commands"). The algorithm gives a score from four operations' cost: insertions, substitutions, deletions, and transpositions. In our study, we adapted a modified algorithm firstly developed for a French orthographical corrector tool based on the ORTHOTEL corpus [35]. The idea of the modification was to weight the "operated components" themselves instead of the "operation". The algorithm itself is written with straightforward rules, where X is an operated cue from a reference command and Y a particular morpho-syntactical variation of this operated cue from the elderly's command:
(X) --> (Y) Cost

After the analysis of the rules, the aim of this algorithm is to give a unique score, which can illustrate the ongoing socio-affective glue statement. This algorithm toolkit is available on open source, using a Pascal compiler (access:

https://github.com/FredericAMAN/GrammaticalDamerauLevenshtein.git). In the EEE corpus, it is the operated objects that seemed to change the glue. For instance, the same operation (addition) can have different effects if we observe the following sentences:

- 1) “here, turn on the light” (‘‘here’’ added)
- 2) “we turn on the light” (‘‘we’’ added)

On a technical aspect, both concern the same automation command, so the actuators must turn on the light. But on a socio-affective point of view, “here” and “we” have got a different value, independently of their linguistic functions or semantics. For instance, 2) illustrated a higher glue statement through the old people's auto-annotations than 1). Moreover, according to the glue behaviour observed in section 3.3 and the error recovery phenomenon, it might be relevant to have a robot feedback and a Smart Home automation activity incrementally produced by an iterative glue assessment. While the robot simulated to fail the recognition on the expected command by executing a “false positive”, on the one hand, the elderly rejected the technology in case 1) as it was a parameter boosting the interaction to increase the glue in case 2). At the same time, instead of giving a similar socio-affective value feedback for both situations, the command activation could be stopped, to avoid the risk to decrease the impression to deal with an intelligent system. The robot could give sounds carrying the idea of its difficulties to answer correctly to the request or provide another stronger glueing sounds to increase the impression a socio-affectively intelligent system. No matter which feedbacks, every decision needs to be very cautiously and collectively discussed as the impact of these sounds (which are not even words having much stronger effects) on the user and the interaction is highly manipulative in this HRI situated context.

PERSPECTIVES: SASI - HRI DIALOG SYSTEM

This first attempt of an ASR prototype aim was to recognise the “relation” between interlocutors, more than the delivered “message” by using multiple LMs describing the different states of socio-affective glue. The primary results are encouraging in perspective to integrate this ASR system into an incremental and iterative global dialogue system - SASI. The present study has only focused on a limited part of the EEE corpus and oral speech recognition. In the long term, the idea would be to connect every multimodal parameter, each one dealing with the processing module dedicated to it, but communicating and influencing the statement of other modules, in a permanent modification. Our proposition to allow this fusion and convergence is to base on a unique “socio-affective glue” score calculated from the features decoded by each module as presented in the modified Grammatical Damereau-Levenshtein algorithm. The module carrying this algorithm, associated with a working memory could then adapt its calculation by comparing the previous statements and possibly predict others. This interconnection gives the temporal organisation of the interactional modules as well as the Smart Home’s context data flow, also changing our way to communicate. All the challenge will reside in knowing when to change the robot’s multimodal socio-affective cues, and how relevantly it will operate the Smart Home, as it is its existence reason carried by its social role.

ACKNOWLEDGMENTS

The LabEx PERSYVAL-Lab (ANR- 11-LABX-0025-01) partially supported this work, as French grants Interabot, parts of BGLE n°2 Investissements d’Avenir. We would like to thank the Awabot company (robotics), Bien A la Maison company (caregiving services), Roger Meffreys elderly housing and the AAPPUI healthcare association for their collaboration. Our thanks to Liliya Tsvetanova, Romain Magnani, Cécile Cottier, Noémie Lagier, Maxence Girard-Rivier, Natacha Borel, Nicolas Bonnefond, Sylvie Humblot and the Getalp members who actively participated in this work’s discussions.

REFERENCES

1. J. Edlund, J. Gustafson, M. Heldner, and A. Hjalmarsson, *Towards human-like spoken dialogue systems*, Speech Communication, vol. 50, no 8, p. 630–645, 2008.
2. R. Rosenfeld and T. Harris, *A Universal Speech Interface for Appliances*, Comput. Sci. Dep., 2004.
3. J. L. Austin, *How to do things with words*. Oxford university press, 1975.
4. J. R. Searle, *What Is an Intentional State?*, Mind, vol. 88, no 349, p. 74, 92, 1979.
5. M. G. Core and J. Allen, *Coding dialogs with the DAMSL annotation scheme »*, AAAI fall symposium on communicative action in humans and machines, vol. 56, 1997.
6. E. A. Schegloff and H. Sacks, *Opening up closing*, Semiotica, vol. 8, no 4, p. 289–327, 1973.
7. G. Aist et al., *Incremental dialogue system faster than and preferred to its nonincremental counterpart*. In Proceedings of the Annual Meeting of the Cognitive Science Society, 29, 2007.
8. R. López-Cózar, Z. Callejas, D. Griol, and J. F. Quesada, *Review of spoken dialogue systems*, Loquens, vol. 1, no 2, p. 012, 2014.
9. M. Paetzel, R. Manuvinakurike, and D. DeVault, *“So, which one is it?” The effect of alternative incremental architectures in a high-performance game-playing agent*, 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, p. 77, 2015.
10. D. Schlangen and G. Skantze, *A general, abstract model of incremental dialogue processing*, in Proceedings of the 12th Conference of the

European Chapter of the Association for Computational Linguistics, p. 710–718, 2009.

11. T. Baumann, M. Atterer, and D. Schlangen, *Assessing and improving the performance of speech recognition for incremental systems*, Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, p. 380–388, 2009.
12. R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, vol. 1. MIT press Cambridge, 1998.
13. O. Lemon and O. Pietquin, *Machine learning for spoken dialogue systems*, European Conference on Speech Communication and Technologies, p. 2685–2688, 2007.
14. J. Schatzmann, K. Weilhammer, M. Stuttle, and S. Young, *A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies*, Knowl. Eng. Rev., vol. 21, no 2, p. 97–126, 2006.
15. O. Pietquin and H. Hastie, *A survey on metrics for the evaluation of user simulations*, Knowl. Eng. Rev., vol. 28, no 01, p. 59–73, 2013.
16. E. Levin and R. Pieraccini, *A stochastic model of computer-human interaction for learning dialogue strategies*, Eurospeech, vol. 97, p. 1883–1886, 1997.
17. N. Audibert, V. Aubergé, and A. Rilliard, *Prosodic Correlates of Acted vs. Spontaneous Discrimination of Expressive Speech: A Pilot Study*, Proc. 5th International Conference on Speech Prosody, Chicago, Etats-Unis, 2010.
18. A.Vanpé and V. Aubergé, *Early meaning before the phonemes concatenation? Prosodic cues for Feeling of Thinking*, GSCP, 2012.
19. Y. Sasa, V. Aubergé, and A. Rilliard, *Social micro-expressions within Japanese-French contrast* », WACAI 2012 Workshop Affect, Compagnon Artificiel, Interaction, Grenoble, 2013.
20. F. Ameka, *Interjections: The universal yet neglected part of speech*, J. Pragmat., vol. 18, no 2, p. 101–118, 1992.
21. Poggi, *The language of interjections », in Multimodal Signals: Cognitive and Algorithmic Issues*, Springer, p.170-186, 2009.
22. M. Schröder, D. K. J. Heylen, and I. Poggi, *Perception of non-verbal emotional listener feedback*, Speech Prosody, 2006.
23. N. Audibert, V. Aubergé, and A. Rilliard, *Acted vs. spontaneous expressive speech: perception with inter-individual variability*, Proc. 2nd International Workshop on Corpora for Research on Emotion and Affect, Marrakech, Maroc, 2008.
24. B. Schuller and A. Batliner, *Computational paralinguistics: emotion, affect and personality in speech and language processing*. United Kingdom: Wiley, 2013.
25. Y. Sagisaka, N. Campbell, and N. Higuchi, *Computing prosody: computational models for processing spontaneous speech*. Springer Science & Business Media, 2012.
26. L. Morency, *Modeling human communication dynamics [social sciences]*, Signal Process. Mag. IEEE, vol. 27, no 5, p. 112–116, 2010.
27. V. Aubergé, Y. Sasa, T. Robert, N. Bonnefond, et B. Meillon, *Emoz: a wizard of Oz for emerging the socio-affective glue with a non humanoid companion robot*, Workshop on Affective Social Speech Signals, Grenoble, France, 2013.
28. V. Aubergé, N. Audibert, and A. Rilliard, « E-*Wiz: a Trapper Protocol for Hunting the Expressive Speech Corpora in Lab*, LREC, 2004.
29. R. Signorello, V. Aubergé, A. Vanpé, L. Granjon, and N. Audibert, *À la recherche d'indices de culture et/ou de langue dans les micro-événements audio-visuels de l'interaction face à face* », WACA, p. 69–76, 2010.
30. G. D. Biasi, V. Auberge, and L. Granjon, *Perception of social affects from non lexical sounds*, GSCP, Belo Horizonte, 2012.
31. V. Auberge, *Attitude vs. emotion: a question of voluntary vs. involuntary control, keynote talk*, in GSCP, Belo Horizonte, 2012.
32. Y. Sasa and V. Aubergé, *Socio-affective interactions between a companion robot and elderly in a Smart Home context: prosody as the main vector of the "socio-affective glue*, Speech Prosody 7, Dublin, Ireland, 2014.
33. V. Aubergé et al., *The EEE corpus: socio-affective "glue" cues in elderly-robot interactions in a Smart Home with the Emoz platform*, 5th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data, Reykjavik, Iceland, 2014
34. Povey, D., et al. (2011). *The Kaldi speech recognition toolkit*. Workshop on automatic speech recognition and understanding. IEEE Signal Processing Society.
35. V. Aubergé, N. Ghneim, and R. Belrhali, *Analyse du corpus Orthotel: apport du traitement automatique à la classification des déviations orthographiques*, Lang. Fr., p. 90–103, 1999.