Enabling Reinforcement Learning for Open Dialogue Systems through Speech Stress Detection

Simon Worgan and **Roger Moore** [1]

Abstract. The human speech signal contains a wide range of paralinguistic information and the interpretation and exploitation of this information presents a fascinating challenge. This paper seeks to demonstrate that emotion is central to the construction of an open-ended dialogue system. Emotion forms a useful, practical, metric that enables a agent to both maintain a user's positive emotional state and allow it to judge and refine its current dialogue strategy. One way to determine the emotional state of the user is through an interpretation of the speech signal. Accordingly, this paper details a method of detecting emotion within speech and then exploiting this information to drive an open-ended dialogue system. After presenting the outline of an advanced open-ended dialogue system we test an initial model to judge the validity of this approach. This simplified model perceives a number of real speech utterances with varying emotional content (ranging from stressed to happy) and learns to manipulate the emotional state of an artificial user through reinforcement learning. The agent acquires the users emotional response to its replies and is able, through a balancing of exploration and exploitation, to maintain the users positive disposition. This initial work is encouraging and clearly shows that emotion can motivate open-ended dialogue. However, a number of substantial challenges remain, including the perception of natural, as opposed to acted, speech and the development of an unsupervised learning approach.

1 Introduction

This paper details how the detection of a users stress levels can be exploited to form a persistent reinforcement learning goal for an open-ended dialogue system. These stress levels will be determined from the paralinguistic features of the users utterances and a simplified model will be constructed to demonstrate the feasibility of this approach. Building on this implemented proof of concept model we will outline how this work can be expanded to provide a rich emotional representation and motivation for the Companions project [27] in future work. This is necessary as the companions project seeks to construct an agent capable of open-ended conversation with a specific user.

Traditionally, spoken dialog systems have been able to learn and reason over uncertainty through a partially observable Markov decision process (POMDP), as shown in figure 1, using reinforcement learning to modify its strategy according to some reward function [29]. POMDP's attempt to address "the problem of choosing optimal actions in partially observable stochastic domains" [13, p. 100] but in practice they frequently prove to be computationally intractable. Accordingly, many researchers settle for a Markov decision process augmented by "a compression of the current belief state"[24, p. 94], to ensure that actions are taken in an uncertain environment in reasonable time.

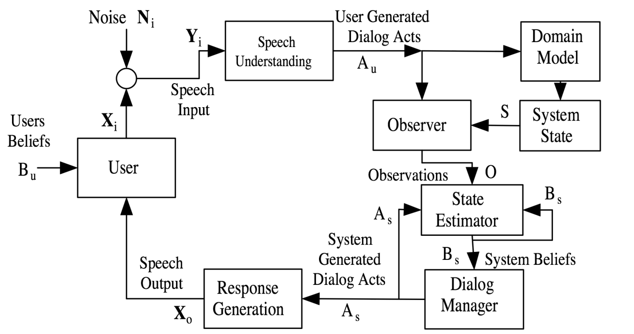


Figure 1. The overall POMDP framework, in this paper the system beliefs, B_s, have been modified to represent a belief about the user's emotional state. These are formed through an acquired observation, O, of the emotional content of the utterance.

This compressed belief state, B_s, drives the reinforcement learning process and its success is entirely dependent upon the derived reward function. In restricted domains with a clearly defined goal (e.g. placing a pizza order) reinforcement learning proceeds by offering a large reward when the task is successfully completed. In the open-ended scenarios proposed by the Companions project how do we redefine the reward function to re-enable reinforcement learning? A persistent conversational agent like the Companion has no clearly defined end state and, currently, no objective measure of a successful interaction. We propose that the Companion should be tasked with maintaining the user's positive emotional state, ensuring continued user satisfaction. Given this goal, emotion becomes a central motivating factor requiring an internal representation of the emotional state of both itself and its user, formed from uncertain knowledge.

This paper will be composed of two parts: a proposed dialogue model that should enable an agent to operate intelligently within a continuous emotional space and an implemented simplification demonstrating the feasibility and possibility of this proposal. The model will be constructed to test the underlying assumptions of the more complex proposal, demonstrating an initial first step that acknowledges the technical challenges [10, 5, 11, 8] remaining within this field. We believe that this implementation demonstrates how various, future, technical improvements can be exploited by a complete dialogue model.

In section 2 we will discuss the challenges of maintaining open-ended dialogue and how this relates to the Companions project. As

[1] Department of Computer Science, University of Sheffield, S1 4DP, UK, email: {simon,r.k.moore}@dcs.shef.ac.uk

part of this discussion we will propose a model that has the potential to navigate the full range of emotional states and adjust its dialogue strategy accordingly. As a first step towards this complex model we will outline an initial prototype in section 3, giving details of a system that can learn and manipulate the simple emotional states of an artificial user. The results of this system will then be discussed in section 4 and the shortcomings and future work highlighted in section 5.

2 Companions project

In previous work [29], Figure 1, reinforcement learning proceeded by forming a belief, B_s, about the current progress towards the successful end of the conversation and assigned an appropriate reward. We have modified this approach, as the belief state is now an understanding of the current emotional state of the user and the reward function is modified accordingly.

At present, the Companion system [12] proceeds through a series of tests and actions arranged into dialogue action forms (DAFS), as the dialogue proceeds DAFS are pushed onto or popped from a conversational stack, allowing the dialogue with the user to proceed. In our own research this strategy will remain; instead of using a POMDP system to learn the specific dialogue strategy we will use reinforcement learning to acquire an intelligent movement through a partitioned emotional 'DAFS space'. This effectively decreases the resolution of the POMDP system, allowing it to learn an overall strategy leaving the details to the existing DAFS. So for example, the sadness of a given situation could be reflected by an empathetic tone of voice and conciliatory paralinguistic features but the dialogue itself remains the providence of the DAFS.

When constructing this space we considered two proposed representations of emotion, the discrete theory of basic emotion [7] and the continuous theory, which maps a range of emotions onto a two dimensional space [28, 4]. We have settled upon the continuous theory (see Figure 2): as instead of arbitrarily defining a number of discrete emotional categories we can attempt a data driven approach, ultimately allowing the Companion to define its own emotional attractors.

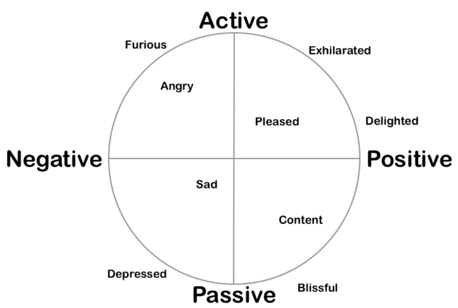


Figure 2. Example of a continuous emotional space, allowing the Companion to define its own emotional attractors.

Before deploying a complete representation of this continuous emotional space we will present a 'proof of concept' model that responds to the stress levels of the user and acquires a mapping between its actions and the resulting emotional state of the user. This decision was taken as despite some success in the classification of large emotional corpora [26, 23, 18, 2] an accurate and rapid understanding of the users' current emotional state will require further research. Building initially on the work of others [22, 5] we will reduce the proposed emotional space to a single dimension representing the current stress level. By detecting key features of stressful speech (e.g. high mean F0, wide pitch range, high energy and fast tempo) it will then be the task of the Companion to move this representation of the user along the continuum towards an unstressed state. Initially, this representation is produced by an artificial user, one which possesses a complex mapping between the Companions' actions and its own emotional state. The task of the Companion is to learn this mapping, through an online reinforcement learning procedure, and to exploit this knowledge to move the user along the simple speech stress continuum into a 'contented' state. The challenges of perceiving the user's emotional state will be simulated by having the artificial user select an utterance from a previously annotated speech stress database [25]. This selection will accurately reflect the user's current emotional condition and by using real speech we have captured the uncertainty inherent in this method. Testing and training then proceeds through multiple interactions between the Companion and the randomly generated artificial user.

Ultimately, this speech stress continuum will be expanded to a full emotional space, mapping Cichosz's reduced feature space [3], which defines a number of emotional speech indicators, to the emotional space defined by Wundt [28]. Over this space three things will operate, a partitioning of the space, and two points representing the user and Companion. A trained self organising map (SOM) [15] will partition the space, defining key 'emotional attractors', it to these attractors that distinct emotionally driven actions will be attached reflecting and expressing the current emotional state of the Companion. Secondly, a representation of the user's emotional state will be defined from a number of multi-modal inputs (paralinguistic features and semantic content), these will form a compressed belief state for the Companion allowing the augmented MDP to proceed. Finally the Companion will maintain a representation of its own emotional state, this will move in response to the user attempting to manipulate them into a positive emotional state, Figure 3. Given this representation the Companion can now learn, through reinforcement learning, how best to alter its own emotions and motivation.

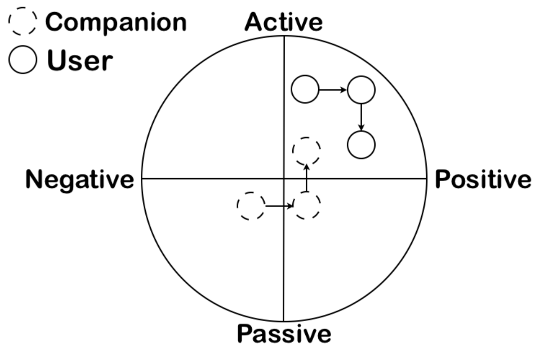


Figure 3. Movement through the continuous emotional space, an understanding of the users location in emotional space allows the companion to respond intelligently by modifying its own location.

By creating a SOM [15], capturing the variability of emotional speech, we can train the system on a wide range of emotional utterances forming distinct attractors within the defined emotional space. Having trained the system these attractors will be tied to a subset of dialogue acts, reflecting the agents current emotional state. Reinforcement learning can then proceed by allowing the Companion to learn how its own movement in this space affects the user. However, this is only possible if the Companion can place a representation of the state of the user into this space, accordingly we need to ascertain the user's current emotional disposition. Clearly, as the model increases in complexity the representation of the emotional state of the user will become increasing subtle and uncertain.

We propose to reduce this uncertainty by extracting emotional information from a number of inputs. By combining the semantic content and acoustic information of the current utterance the uncertainty of the current emotional state can be reduced [17, 16]. Having established this understanding the Companion will have a number of tools with which to manipulate the situation.

The Companion will converse with the user on a number of levels, the tone and paralinguistic nature of the utterance can be adjusted to convey emotional content, the structure of the dialogue act can be similarly adapted and the overall conversational strategy can be optimised. When exploiting these features it is important to maintain a consistent personality that accurately conveys the capabilities and limitations of the Companion, any gap between user expectation and reality would form an obvious source of dissatisfaction. As shown by previous WOZ studies [20] simple differences to tone of voice or appearance can cause dramatic differences in user behaviour. Accordingly, the Companion will present a unified personality (encompassing appearance, voice, personality and behaviour) which accurately manages the user's expectation, as we have demonstrated emotion lies at the heart of this approach.

3 Initial reinforcement learning

To test the fundamentals of the system proposed in section 2 an initial model will be constructed and analysed. The implemented system will consist of a number of components. The agent itself will contain two stages, a perceptual stage, section 3.1, which determines the emotional content of a speech signal and a reinforcement learning stage, section 3.2, which learns how to maintain the users positive emotional state given this information. The 'user' in this system is entirely artificial and consists of two mappings, one which maps from a current emotional state to an utterance and the other which maps from the agents response to an adjustment of its own emotional state. To successfully maintain the user's positive emotional state the agent needs to acquire these two mappings.

3.1 Perception of emotion from speech

To provide the artificial users utterances training and test data was obtained from a database of German emotional speech [1] consisting of 10 native German speakers (5 female and 5 male) simulating 6 basic emotions with 10 utterances per emotion. To simplify our training task we selected 2 basic emotions (stress and joy) and an individual neural network was trained on the F0 values [14] extracted from each utterance.

The F0 values corresponds to the central frequency of speech and is believed, among other things, to convey paralinguistic qualities of speech [6], it has been hypothesised that this region contains a large amount of emotional content in speech. By training each network on a range of utterances from an individual speaker and stripping out the linguistic content of the utterance we are able to reduce the variance of the task and focus upon its emotional content.

The first stage of learning then proceeds through Levenberg-Marqundt [19] backpropogation, for each utterance 204 F0 values where extracted from 500ms of speech. After supervised training, figure 4(a), on the acted emotional speech the agent proceeds to the second stage, where it learns to exploit this information through conversation with the artificial user.

3.2 Use of emotion in dialogue

Having acquired a connection between speech and emotion the agent will try to manipulate the user into a positive emotional state. The user responds by moving along an abstract emotional continuum ranging from upset to happy (-1 to 1). Attractors along this scale at intervals of 0.2 capture how a users current emotional state influence their response to utterances, i.e., a certain phrase will not produce the same response in all scenarios. These attractors capture the behaviour of the converged SOM that will be present in the final system. Accordingly, to successfully maintain the users positive state the system needs to learn the mapping between speech and emotional consequence at each attractor point.

The artificial user conveys its current emotional state by selecting an appropriate speech signal from the German speech database. It then listens to the systems response and adjusts its emotional state according to its mapping between response and emotion, modified by a sensitivity parameter, σ. This parameter represents a proportion between the user's current state and the new state proposed by its mapping.

The agent's understanding of the users selected response forms a belief, B_s, about the users current emotional state. This belief is then modified by the agents learning rate, λ, to update the agents understanding of the user's mapping from perceived utterance to emotional state. Here, λ represents a proportion between the systems current understating and B_s. This cycle of user utterance and agent reply continues until is is clear that the users emotional state has converged. Each round of utterance and reply is described as one time step in section 3.3.

During reinforcement learning we have modified the standard exploration approach α [29] buy adding the variable ρ, as shown in equation 1. This variable is the euclidean distance to the desired emotional state of the user. Consequently, as the user is perceived to approach this state the agent feels increasingly at liberty to explore the space of possible utterances. As defined in equation 1, when combined with a variable decay rate, δ, the system avoids premature convergence on a local optimum and balances exploration/exploitation in response to the user, i.e. when the user is unhappy it will focus on making them happy, when they are happy it can indulge in exploration of the space of possible mappings.

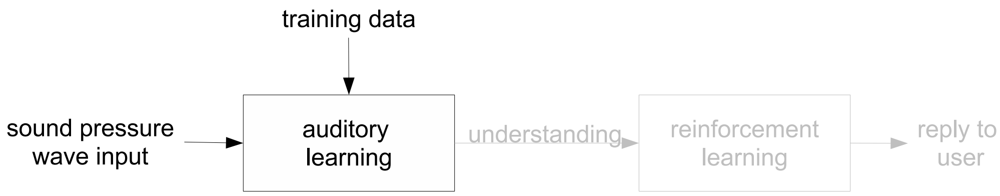
$$\alpha = E^{(-\frac{\rho+1}{\delta}x)} \tag{1}$$

In equation 1 α defines a proportion of the subset of possible utterances, attached to an emotional state, of which one is selected at random, allowing for exploration around the optimum utterance.

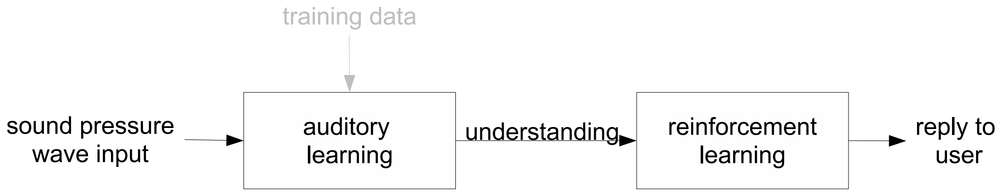
3.3 Experimental work

Having developed this model we will now investigate whether it can learn the users emotional responses and maintain the users positive emotional state. We will also investigate the systems robustness to parameter variation (δ,λ,σ) and the effect of a variable balance between exploration and exploitation, captured by ρ.

In all experiments we will record the results from 5 different German speakers (2 female, 3 male) averaged over 20 runs. We will be recording the changing emotional state of the user and the average



(a) Stage one training, supervised learning of a mapping between speech and emotion.



(b) Stage two training, reinforcement learning proceeds through dialogue.

Figure 4. The two training stages of the agent as it attempts to learn the mapping between its own utterances and the user's emotional state and attempts to manipulate the user into a favorable emotional state.

error for the system's understanding of the user's signal to emotion mapping.

4 Results

During supervised training, figure 4(a), the average classification error (stressed speech $= -1$/happy speech $= 1$) over 100 trials was 0.0049 with a standard deviation of 0.018. This suggests that for our limited purposes the change in F0 is sufficient to distinguish between these two types of emotional speech.

The results of stage 2 show in, figure 5, that the system is highly robust to parameter variation, after 800 time steps the systems understanding had converged to a solution and the users emotional state was maintained. As can be seen in figure 5(b) without any learning taking place, the control experiment, the agent fails to establish a positive emotional state in the user as it cannot correct its initially random signal to emotion mapping.

Under a reasonable parameter set ($\delta = 500$, $\lambda = 0.5$ and $\sigma = 0.2$) without variable exploration the system reduces the absolute average error to 0.81 with variable exploration the error is reduced to 0.67. Averaged over 100 runs (20×5 speakers) and between 95 to 125 signal to emotion mappings, depending on the chosen German speaker, the resulting p-value is less than $1\mathrm{E}^{-51}$ establishing the clear significance of exploration.

5 Conclusion

This paper has presented an implementation of an open ended dialogue strategy, through the implementation of the variable exploration/exploitation balance the agent can continue to learn and function in response to the users emotional state. However, it is important to emphasise the abstract nature of this initial model, even though we use real speech our user is entirely artificial and, as figure 5 shows, it can be argued that it doesn't present enough of a challenge for our system. However, we believe that this model establishes the plausibility of this approach and, as shown in figure 5(b), without learning the system fails.

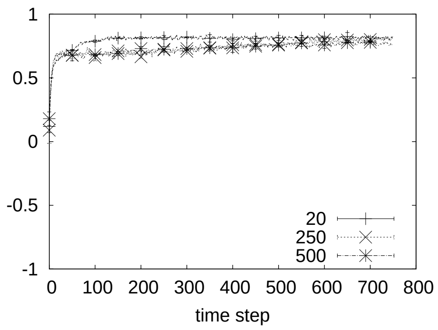
Due to the abstract nature of the user it becomes difficult to demonstrate the advantages of the variable exploration parameter, ρ. As shown by figure 5(d) the emotional state of the user does not suffer when this parameter is removed. However, the resulting difference in the average learning error demonstrates the importance of this parameter and we believe that as increasingly complex and realistic models are constructed its role will become increasingly important.

In previous work [21, 9] it has been shown that natural speech, with emotional ambiguity and few incidents of 'full-blown' emotion, is significantly harder to learn than acted speech. This has implications for our initial learning stage, Figure 4(a), which is further complicated by the fact that the Companion needs to make sense of emotional speech from the outset. In typical use a comprehensive, annotated, dataset of emotional speech tailored to each user will be unavailable, making a supervised learning approach impossible. Additionally, attempts to build a general emotional speech recogniser were unsuccessful as the variability between speakers was too great for the neural network, obscuring the relevant emotional features. Clearly, an advanced unsupervised approach will have to be developed to recognise real emotion in speech.

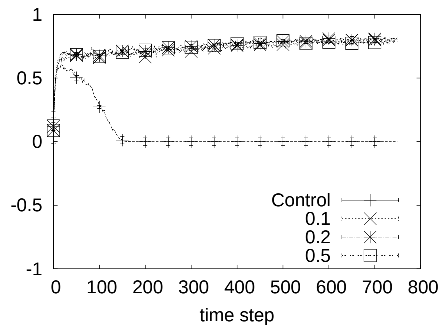
Although the results remain robust at no point do they reach an optimum emotional state, we believe that this is due to the perceptual error inherent when working with real speech. As a result misclassification errors occur preventing the system from achieving an optimum result. Additionally, as the system moves towards an optimum it is at increasing liberty to explore a range of utterances, accordingly the system trades a perfect emotional state for the opportunity to acquire a range of possible mappings.

Clearly, a number of challenges remain but this initial model has served its purpose. We have demonstrated that emotional speech can motivate an open-ended dialogue and developed a method of balancing the twin goals of maintaining a positive emotional state while learning a users response to emotional utterances. We believe that as the technical challenges, highlighted in this paper, are overcome this approach will become increasingly important, eventually reaching the point where it can converse with real users.

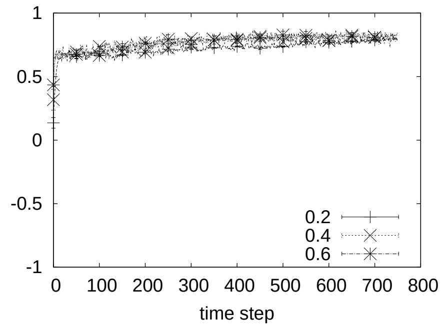
In this paper we have detailed two things: an implemented proof of concept model demonstrating how the emotion of a dialogue can enable the application of reinforcement learning in an open-ended conversational system and a plan detailing how this work can be expanded to capture and exploit a wide range of uncertain human emotions. In future work we will begin to address this uncertainty by



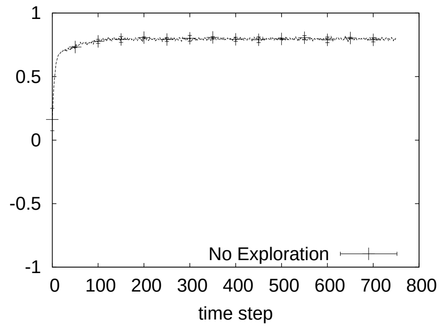
(a) Robustness to a varying decay parameter, δ



(b) Robustness to a varying learning parameter, λ. Without learning the system fails.



(c) Robustness of the system to varying user sensitivity, σ.



(d) Without flexible exploration the system can still maintain a positive emotional state.

Figure 5. Demonstrating a robustness to parameter variation, and the abstract nature of the user, the system is able to maintain the users positive emotional state (y-axis>0) under a variety of scenarios. Results obtained from an average of 5 speakers with 20 runs per speaker, error bars have been plotted but are too small to show.

proposing the exploitation of multiple sources of information, integrating speech, language and the environment. Ultimately, we hope to present a unified Companion, one which can accurately assess and respond to the user's emotional state.

6 Acknowledgement

This work was funded by the Companions project (www.companions-project.org) sponsored by the European Commission as part of the Information Society Technologies (IST) programme under EC grant number IST-FP6-034434.

REFERENCES

[1] F Burkhardt, A Paeschke, M Rolfes, and W Sendlmeier, 'A database of german emotional speech', in *Ninth European Conference on Speech Communication and Technology*, Lisbon, Portugal, (2005).

[2] S Casale, A Russo, and S Serrano, 'Multistyle classification of speech under stress using feature subset selection based on genetic algorithms', *Speech Communication*, **49**(10-11), 801–810, (2007).

[3] J Cichosz and K Slot, 'Low-dimensional feature space derivation for emotion recognition', in *Ninth European Conference on Speech Communication and Technology*, Lisbon, Portugal, (2005).

[4] R Cowie, E Douglas-Cowie, S Savvidou, and E McMahon, ''feeltrace': An instrument for recording perceived emotion in real time', in *ISCA Tutorial and Research Workshop on Speech and Emotion*, Newcastle, Northern Ireland, UK, (2000).

[5] R Cowie, E Douglas-Cowie, N Tsapatsoulis, G Votsis, S Kollias, W Fellenz, and JG Taylor, 'Emotion recognition in human-computer interaction', *Signal Processing Magazine, IEEE*, **18**(1), 32–80, (2001).

[6] D Crystal, *A first dictionary of lingusitics and phonetics*, WileyBlackwell, 1980.

[7] P Ekman, 'Basic emotions', in *Handbook of Cognition and Emotion*, Wiley, (1999).

[8] Barreto et al., 'Non-intrusive physiological monitoring for automated stress detection in human-computer interaction', *Human Computer Interaction*, **4796**, 29–38, (2007).

[9] P. Greasley, J. Setter, M. Waterman, C. Sherrard, P. Roach, S. Arnfield, and D. Horton, 'Representation of prosodic and emotional features in a spoken language database', in *13th International Congress of Phonetic Sciences*, pp. 242–245, Stockholm,Sweeden, (1995).

[10] P Greasley, C Sherrard, and M Waterman, 'Emotion in language and speech: Methodological issues in naturalistic approaches', *Language and Speech*, **43**(4), 355–375, (2000).

[11] CS Hopkins, RJ Ratley, DS Benincasa, and JJ Grieco, 'Evaluation of voice stress analysis technology', *38th Annual Hawaii International Conference on System Sciences*, 1–10, (2005).

[12] http://www.companions project.org/.

[13] LP Kaelbling, ML Littman, and AR Cassandra, 'Planning and acting in partially observable stochastic domains', *Artificial Intelligence*, **101**(1-2), 99–134, (1998).

[14] S Kim, PG Georgiou, S Lee, and S Narayanan, 'Real-time emotion detection system using speech: Multi-modal fusion of different timescale features', in *IEEE 9th Workshop on Multimedia Signal Processing*, pp. 48–51, (2007).

[15] T Kohonen, 'The self-organising map', *Proceedings of the IEEE*, **78**(9), 1464–1480, (1990).

[16] S Lauria, 'Talking to machines: Introducing robot perception to resolve speech recognition uncertainties', *Circuits Systems Signal Processing*, **26**(4), 513–526, (2007).

[17] C Lee, S Narayanan, and R Pieraccini, 'Combining acoustic and language information for emotion recognition', in *Seventh International Conference on Spoken Language Processing*, Denver, CO, (2002).

[18] YL Lin and G Wei, 'Speech emotion recognition based on hmm and svm', in *International Conference on Machine Learning and Cybernetics*, volume 8, (2005).

[19] D. W. Marquardt, 'An algorithm for least squares estimation of nonlinear parameters', *Journal of the Society of Industrial and applied Mathematics*, **11**(2), 431–441, (1963).

[20] R Moore and A Morris, 'Experiences collecting genuine spoken enquiries using woz techniques', in *Fifth DARPA Workshop on Speech & Natural Language*, Narriman, NY, (1992).

[21] P-Y. Oudeyer, 'The production and recognition of emotions in speech: features and algorithms', *International Journal Of Human-Computer Studies*, (2003).

[22] A Paeschke, 'Global trend of fundamental frequency in emotional speech', in *Speech Prosody*, Nara, Japan, (2004).

[23] CH Park and KB Sim, 'Emotion recognition and acoustic analysis from speech signal', in *International Joint Conference on Neural Networks*, (2003).

[24] N Roy, J Pineau, and S Thrun, 'Spoken dialogue management using probabilistic reasoning', *38th Annual Meeting of Association for Computational Lingusitics*, (2000).

[25] S Scherer, F Schwenker, and G Palm, 'Emotion recognition from speech using multi-classifier systems and rbf-ensembles', in *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks*, 49–70, Springer, Berlin, (2008).

[26] B Schuller, G Rigoll, and M Lang, 'Hidden markov model-based speech emotion recognition', in *International Conference on Multimedia and Expo*, volume 2, pp. 1–4, (2003).

[27] Y Wilks, 'Is there progress on talking sensibly to machines?', *Science*, **318**(5852), 927, (2007).

[28] W Wundt, *Grundriss der Psychologie*, A. Kroner, 1913.

[29] S Young, 'Talking to machines (statistically speaking)', in *Seventh International Conference on Spoken Language Processing*, Denver, CO, (2002).