

# **Human-computer interaction systems and elderly users: Linguistic specifications for prosodic modelling**

Christina Alexandris  
Department of German, University of Athens, Greece

<https://doi.org/10.36505/ExLing-2010/03/0002/000122>

## **Abstract**

The present linguistic specifications and language-independent parameters for prosodic modelling aim to provide a general framework for facilitating both the construction and the evaluation processes of the Dialog Interface of Human-Computer Interaction Systems in Service for the Elderly and the related prosodic modelling. The proposed specifications target to the features of Comprehensibility and User-friendliness in the spoken output and the overall safety, efficiency and reliability of the System's performance.

Key words: comprehensibility, user friendliness, prosodic emphasis, speed rate

## **Data and user requirements**

General parameters as usable, transferable and language independent specifications are presented here for the Speech Processing Module and Dialog Interface of Human- Computer Interaction Systems in Service for the Elderly. These parameters aim to contribute to the achievement of Comprehensibility and User-friendliness adapted to the needs of the Elderly Users. With the proposed approach, we aim to limit empirical prosodic modelling and case-specific management of a System's spoken output and to provide a general framework for facilitating the construction of the Dialog Interface and the related prosodic modelling for the Elderly as the user group. Here, we focus on experience and data gained from European Union projects (European countries including the Mediterranean and Northern Europe, language and culture-specific factors), namely data from User Requirements Analysis in Work Packages from the respective projects, recorded data, questionnaires, dialog modelling corpora and studio recordings from the respective projects, namely: the SOPRANO Project (involving smart environments and services for the Elderly, the HEARCOM Project (speech technology applications for Users with hearing problems) and the CitizenShield Dialog System for consumer complaints (Greek national project), with senior citizens constituting a significant percentage of the System's users ([www.polias.gr](http://www.polias.gr)). From a psychological perspective, previous studies in the case of Systems involving Human Computer Interaction (HCI), have demonstrated that the User is more likely to use the System and oversee minor flaws if the System is user-friendly and its spoken output is characterized by naturalness and

in some cases may create “an attachment emotion” towards the system (Matsumoto et al., 2009). In this case, the Elderly User is also more likely to make the effort and gain the most benefit from the System’s existing capabilities.

Prosodic modelling constitutes one of the typical strategies used in Dialog Systems (Kellner, 2004), contributing to the task efficiency and service efficiency of the system (Moeller, 2005). In Task-related applications, for the General Public, prosodic modelling is observed to contribute both to the achievement of clarity, lack of ambiguity and user-friendly style (Alexandris, 2009). For the requirements of the Elderly as a User-Group, these targets are summarized as “Comprehensibility” (I) and “User-Friendliness” (II). Here, prosodic modelling is designed within a Speech Act framework. Prosodic modifications are related to paralinguistic elements (prosodic emphasis, pause-length) and pragmatic elements (speech act differentiation). The main features involved in the proposed prosodic modelling are (A) prosodic emphasis, (B) speed, (C) amplitude and (D) overall tone of utterance.

### **Prosodic emphasis**

Prosodic modelling of all utterances related to the Speech Acts for Task-oriented Dialogs is based on the use of prosodic emphasis (A) on the sublanguage-specific elements constituting the most important information in the sentence’s semantic content, as well as sublanguage-independent elements, such as negations and elements expressing time, space (movement), quality and quantity (such as “now”, “over”, “red (button)” and “two (minutes)”), used for the achievement of Precision, while prosodic emphasis on sublanguage-specific expressions and terminology (such as “airconditioner”, “button” and “turned off”) is used for the achievement of the Comprehensibility (I) resulting to Directness, (Alexandris, 2009). From a prosodic aspect, it is observed that User-friendliness (II) can be achieved with prosodic emphasis on expressions related to the User-System Relationship (“Usr-Sys-Rel” expressions), especially in Speech Acts labelled here as “Non-Task-related Speech Acts” such as “Apologize” and “Thank”, or following or preceding the Task-Related Speech Act (Heeman et al, 1998), for example, the utterance “To make sure that the house is safe, I will ask you a few more questions” (“Introduce-new-task”) following or preceding the Task-Related Speech Act (“Y/N Question”) based on the User-System Relationship. The “Usr-Sys-Rel” expressions can be subsumed under the general category of expressions involving the System’s or User’s positive intention or cooperation. These words express the system’s intention (action) to serve the user, expressions (usually verbs) indicating the system’s apologies, failure or error in respect to a task executed to serve the user. Usr-Sys-Rel expressions are also observed to be words (usually verbs) expressing the user’s actions or intentions, or words (mainly nouns)

expressing a task related to the actual interaction involving good intentions or a task related to the system's services. Usr-Sys-Rel expressions can be categorized as (1) system-service verbs and nouns (nominalization of verbs) ("cooperation", "give", "show"), (2) system-intention verbs and nouns ("help", "assist") and (3) user-intention (or user intended action) verbs ("wish", "want"). We note that the User-System Relationship is remarkably and evidently expressed in Usr-Sys-Rel verbs in verb-framed and pro-drop languages such as Greek (or, for example, Spanish or Italian), where the emphasized finite verb as a Usr-Sys-Rel expression, contains the features of the verb's subject, in this case the System or the User. Typical examples are the expressions "doso" (=I-give/present), "deixso" (=I-show/present) and "exshypyre'tiso" (=I-give-service/assist).

### **Speed, amplitude and overall tone**

The specific values of the speed rate of the utterances produced in the System's spoken output may vary according to language-type. For the achievement of Comprehensibility (I) and User-friendliness (II) for the Elderly as a user group, as a default parameter, the speed of the spoken output should be lower than the threshold of the average speed of a news broadcast, but higher than the average speed maintained for reading aloud to small children. Another factor contributing to the speed rate is the management of the pauses in utterances produced. A short pause ([Srt-P]) before keywords to be emphasized (and their articles and/or pronouns, if applicable) contributes to the overall speed of the utterances produced. For English, a short pause is also inserted between the components of ACTION-TYPE (Malagardi and Alexandris, 2009) constituting expressions composed of more than one word, such as the expressions "turned-off" and "switched-on", or any elements containing key-information (and receiving prosodic emphasis, here indicated in italics), such as in the example: "OK. The [Srt-P] *air conditioner* is *switched* [Srt-P] *on*". The absolute values of the short pauses may vary according language-type, since they are related to the overall speed rate of the utterances produced. Similarly to the feature of speed (B), the specific values of the overall amplitude (C) of the utterances produced in the System's spoken output may also vary according to language-type. For example, some languages appear to have a characteristically loud form of communication by their speakers, in relation to other languages, especially in Northern Europe, where, in most types of transactions and related Speech Acts, lower amplitude is preferred and is acceptable by most users. However, overall amplitude can be adjusted, according to the User's needs. The acceptable tone (D) of voice in the System's spoken output may also vary according to culture and language-type, constituting a substantial factor in the requirements of the Elderly as a

user group: a characteristically vivid and expressive tone is maintained in some languages (i.e. Italian) while in others languages, a tone associated with reliability and responsibility is usually preferred (i.e. German). In most cases, at least for the languages and their Native Speakers in Europe, developers in Human-Computer Interaction systems have to balance between a tone associated with reliability and responsibility, while at the same time User-friendliness and naturalness must be preserved. We note that fine-tuning in respect to appropriate nuance tone is easier to be accomplished by pre-recorded output produced by a trained speaker than by Speech Synthesis, since, unlike prosodic emphasis, amplitude and speed, tone is both highly language and culture specific and cannot be easily encoded and integrated in the System's Speech Processing Module.

### References

- Alexandris, Ch. 2009. A speech-act oriented approach for user-interactive editing and regulation processes applied in written and spoken technical texts. In Jacko, J.A. (ed.), *Human-Computer Interaction, Part II, HCII 2009*, vol. 2, LNCS 5611, 645–653, Berlin Heidelberg: Springer-Verlag
- Heeman, R., Byron, D., Allen, J.F. 1998. Identifying discourse markers in spoken dialog. *Proc. of the AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, Stanford, USA.
- Kellner, A. 2004. Dialogsysteme. In Carstensen, K.U, Ebert, Ch., Endriss, C., Jekat, S., Klabunde, H. Langer, H. (eds.), *Computerlinguistik und Sprachtechnologie, Eine Einfuehrung*, 532-539, Muenchen: Spektrum Akademischer Verlag.
- Matsumoto, N., Ueda, H., Yamazaki, T., Murai, H. 2009. Life with a robot companion video analysis of 16-days of interaction with a home robot in a “Ubiquitous Home” environment, J.A. Jacko (ed.), *Human-Computer Interaction, Part II, HCII 2009*, LNCS 5611, 341–350, Berlin Heidelberg: Springer-Verlag.
- Malagardi, I., Alexandris, Ch. 2009. Verb processing in spoken commands for household security and appliances. In Stephanidis, C. (ed.), *Universal access in HCI, Part II, HCII 2009*, vol. 6, LNCS 5615, 92–99. Heidelberg: Springer-Verlag.
- Moeller, S. 2005. *Quality of telephone-based spoken dialogue systems*. New York, Springer.
- <http://www.soprano-ip.org/>
- <http://hearcom.eu/main.html>
- [www.polias.gr](http://www.polias.gr)