CABA^{2}L A BLISS PREDICTIVE COMPOSITION ASSISTANT FOR AAC COMMUNICATION SOFTWARE

Nicola Gatti
Dipartimento di Elettronica e Informazione, Politecnico di Milano
Piazza Leonardo da Vinci 32, I-20133, Milano, Italy

Matteo Matteucci
Dipartimento di Elettronica e Informazione, Politecnico di Milano
Piazza Leonardo da Vinci 32, I-20133, Milano, Italy

Keywords: AAC languages, accessibility to disabled users, hidden Markov model, intelligent user interface, symbolic prediction.

Abstract: In order to support the residual communication capabilities of verbal impaired peoples softwares allowing Augmentative and Alternative Communication (AAC) have been developed. AAC communication software aids provide verbal disables with an electronic table of AAC languages (i.e. Bliss, PCS, PIC, etc.) symbols in order to compose messages, exchange them via email, or vocally synthetize them, and so on. A current open issue, in thins kind of software, regards human-computer interaction in verbal impaired people suffering motor disorders. They can adopt only ad-hoc input device, such as buttons or switches, which require an intelligent automatic scansion of the AAC symbols table in order to compose messages. In such perspective we have developed CABA^{2}L an innovative composition assistant exploiting an user linguistic behavior model adopting a semantic/probabilistic approach for predictive Bliss symbols scansion. CABA^{2}L is based on an original discrete implementation of auto-regressive hidden Markov model called DAR-HMM and it is able to predict a list of symbols as the most probable ones according to both the previous selected symbol and the semantic categories associated to the symbols. We have implemented the composition assistant as a component of BLISS2003 an AAC communication software centered on Bliss language and experimentally validated it with both synthetic and real data.

1 INTRODUCTION

Nowadays, millions of verbal impaired people live in the world (Bloomberg and Johnson, 1990); their communication capabilities are permanently or temporarily corrupted and, for this reason, most of them suffer a condition of social exclusion. Verbal impaired people can not adopt canonic communicative media (Fodor, 1983), such as natural language, and, as the clinical experience evidences, their primary need is to try alternative ways, according to their residual capabilities, to communicate. In 1983 the International Society for Augmentative and Alternative Communication (ISAAC, 1983) has been established in USA with the aim to develop alternative instruments to allow verbal impaired people to communicate. ISAAC has been involved in developing both languages, namely, Augmentative and Alternative Communication languages (AAC) (Shane, 1981), and aids in order to support residual communicative capabilities in verbal disables. AAC languages are usually based on symbols and exploit peculiar composition rules simple enough to be learnt and used by verbal disables. Among the AAC languages we can cite: Bliss, PCS, PIC, PICSYMB, CORE, and Rebus.

Currently, disables adopt paper tables (see Figure 1) containing their most used AAC symbols and point in such tables the symbols related to what they want to communicate. In the AAC field other AAC aids exist, such as VOCAs (i.e. smart tablets that associate vocal inputs to specific symbols), but they evidence severe limitation with respect to effective verbal disables needs since they present a limited set di predefined sentences. Given this scenario, information technology plays a relevant role by providing the verbal disabled people with aids, such as AAC software applications, to support their communication. In fat, AAC software applications provide verbal impaired people with an electronic table of symbols where they can select AAC symbols to compose messages adopting canonical or ad-hoc AAC devices (e.g. joystick, tablet, switch, etc.). In addition, they offer other features, such as email message exchange and vocal synthesis.

Gatti N. and Matteucci M. (2004).
CABA2 L A BLISS PREDICTIVE COMPOSITION ASSISTANT FOR AAC COMMUNICATION SOFTWARE.
In *Proceedings of the Sixth International Conference on Enterprise Information Systems*, pages 89-96
DOI: 10.5220/0002606600890096

A current open issue concerning AAC communication software aids regards human-computer interaction in disables with motor disorders (Lee et al., 2001), that represent about the 60% of verbal impaired people. Motor disordered people are not able to use canonical input devices, such as keybord and mouse, but they can use only ad-hoc devices, such as buttons or switches according to their residual motor capabilities. Such devices operate providing the AAC software aid with an on/off input, so, in order to select AAC symbols from a table, it is required an automatic scansion of such table underlining the symbols they can select. The intelligent scansion mechanisms currently adopted in AAC software do not assure a relevant reduction of the time spent by verbal disables to compose messages: in fact person evidencing motor disorders can spend few minutes to compose a simple sentece. In such perspective we have developed CABA^{2}L (Composition Assistant for Bliss Augmentative Alternative Language) an innovative composition assistant that performs a predictive scansion of Bliss symbols and reduces up to 60% the time required to compose a message. CABA^{2}L is based on a discrete implementation of auto-regressive hidden Markov model (Rabiner, 1989) called DAR-HMM and it predicts a list of symbols as the most probable according to both the last selected symbol and the semantic categories associated to symbols. Moreover, the predictive model embedded in CABA^{2}L can be adapted to the specific disable user to better match his/her peculiar linguistic behavior.

The paper is structured as follows. In Section 2, we introduce current scansion mechanisms for verbal disables and underline prediction issues. Section 3 introduces Bliss symbolic prediction issues and probabilistic prediction techniques. Section 4 describes DAR-HMM and its implementation. Section 5 reports the experimental results we have obtained. Finally, Section 6 concludes the paper.

2 SENTENCE COMPOSITION IN PEOPLE WITH MOTOR DISORDERS

Currently some AAC software aids provide people suffering motor disorders with an automatic scansion of the symbol table (see Figure 1). A generic scansion mechanism can be described as follows: an highlight moves autonomously on an AAC symbol table according to a specific strategy, when the requested symbol is highlighted the user can select such symbol activating the device, then the highlight starts to move again. Each scansion mechanism is characterized by a specific ad-hoc input device and the scansion strategy.

Ad-hoc input devices allow the verbal disable to start the scansion, select symbols, close the sentence, and activate features such as vocal synthesis. Each user adopts the device that best matches with his/her residual motor capabilities. The alternative/residual motor capabilities used are usually: blowing with the mouth, closing the eyes, pushing big buttons.

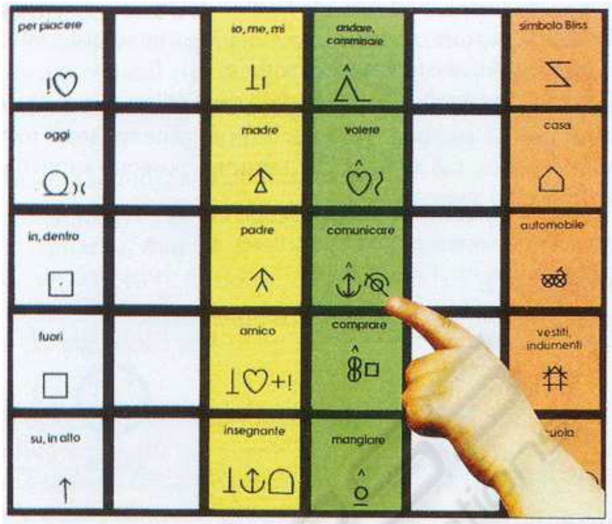

Figure 1: An AAC symbols table

The scansion strategy determines what is the next symbol to highlight. In literature several automatic scansion strategies for AAC communication aids can be retrieved (Higginbotham et al., 1998) and each one of them exhibits advantages and drawbacks. Such strategies can be classified in (Swiffin et al., 1987):

- *linear scansion*: the symbols are presented sequentially from the first symbol of the table to the last one;
- *row-column scansion*: at first the rows are scanned, once the disable has selected a row, its columns are scanned (or vice versa) (Simpson and Koester, 1999);
- *scansion at subgroups*: the symbols are presented in groups fewer and fewer up to only one symbol;
- *predictive scansion*: it predicts the most probable symbols the user will use to continue the sentence according to a model of the user, and it presents the most probable ones.

The choice of the most adequate scansion strategy for the peculiar user depends on several factors, such as the mental and motor residual capabilities, the adopted AAC language and the size of the user symbols table. With respect to the user mental capabilities, both scansion at subgroups and row-column scansion require the user to remember exactly the place of the desired symbol, so they can not be used by disables evidencing severe mental impairments. With respect to the size of the symbol table linear scansion, row-column scansion and scansion at subgroups do not offer good performance if the number of symbols is elevate (50 symbols and more). Hence, current non predictive scansion strategies do not allow to verbal disables suffering motor disorders a relevant reduction in the time spent to compose sentences (Lee et al., 2001).

Although predictive scansion strategies could assure better performance (Koester and Levine, 1994), they are currently adopted in a small number of AAC assistitive technology aids. In particular *symbolic prediction* is currently adopted only in VOCAs, but it evidences severe limitations: it predicts symbols according to a strict set of sentences previously registered and do not exploit a linguistic behavior model of the user. In such a way such prediction system allows the user to compose a fixed number of messages and it is not able to generalize allowing the composition of new messages (Higginbotham et al., 1998). In literature numerous predictive techniques have been developed, but they have been applied mainly in alphabetical prediction. In such context the main prediction techniques (Aliprandi et al., 2003) employ a *statistical* approach (based on hidden Markov models and Bayesian networks), a *syntactic* and *strong syntactic* approach (based on linguistic models), a *semantic* approach (based on semantic networks), and hybrid approaches. To the best of our knowledge, currently symbolic predictive models do not exist.

The main issue with alphabetical predictive techniques, that prevents their use for symbolic prediction, is related to the size of the dictionary of items to be predicted and their composition rules. In fact, alphabetical prediction operates on a limited number (about 20) of items, the alphabetic signs, that can be organized in words known *a priori*. Conversely, symbolic prediction operates on a set of symbols variable in number that can be organized in different sequences according to the peculiar user linguistic capabilities. In addition alphabetical prediction techniques do not match with the symbolic prediction issue. On the other side, a pure statistical approach does not keep into account the peculiar AAC language structure, in fact each verbal impaired user adopts/develops an own syntactic model according to his/her residual mental capacities. This is also the reason for which the utilization of a pure syntactic approach for any user can not be achieved, and a pure semantic approach does not address the variability related to the residual user capacities.

We consider an ad-hoc hybrid approach as the right choice in this context; in the following sections of the paper we focus on the description of this prediction model since it represents the most original part of our work.

3 BLISS SYMBOLS PREDICTION AND GRAPHICAL MODELS

In our work we focus on the Bliss language (Bliss, 1966), since it is the most adopted and expressive among AAC languages. In the design of a composition assistant to predicts Bliss symbols, a set of peculiar requirements regarding both the *human-computer interface* and the *prediction model* can be established.

The composition assistant should suggest a limited number of symbols (around 4-5) not to confuse the disable user (Koester and Levine, 1994), the prediction must be accomplished in real time, and the scansion rate must be adaptable with the user needs (Cronk and Schubert, 1987). This last aspect addresses issues due to the high variability of residual mental and motor capabilities, in fact the composition assistant should be able to adapt the scansion rate according to the time required by the specific disable to read and select the highlighted symbol. With respect to the prediction model, a verbal impaired user can adopt all the Bliss symbols (about 2000), even if he/she usually adopts only a part of them (usually from 6-7 to 200), and it should be taken into account that the symbol to be predicted depends in some extents on the symbols selected previously.

We have adopted a semantic/probabilistic approach to model the user language behavior and we use this model in order to predict the most probable symbols to be suggested by an automatic scansion system. We have used the semantic approach to take advantage of a Bliss symbols categorization and the probabilistic approach both to take into account for uncertainties in the user language model and to give a reasonable estimate of the reliability of the proposed prediction.

In CABA^{2}L we have used a graphical model based on a variation of a classical Hidden Markov Models (HMM). Classical HMMs involve *states* and *symbols*, in particular they relate the probability that a particular symbol is emitted to the probability that the system is in particular state. Moreover they use a stochastic process to define the transition from a state to the other (see Figure 2).

In HMM a particular sequence of observation (i.e. observed symbols) is generated by choosing at time $t = 0$ the initial state $s_i \in S$ according to an initial probability distribution $\pi(0)$, a symbol v_k is generated from a multinomial probability distribution b_k^i associated to state s_i, and the system move from the present state s_i to the next state $s_{i'}$ of the sequence according to a transition probability $a_{ii'}$ to generate the next symbol. States in this model are not directly observable; symbols represent the only information that can be observed, and this is the reason for the term *hidden* in the model name. Notice that classical HMMs consider symbols as independent from each

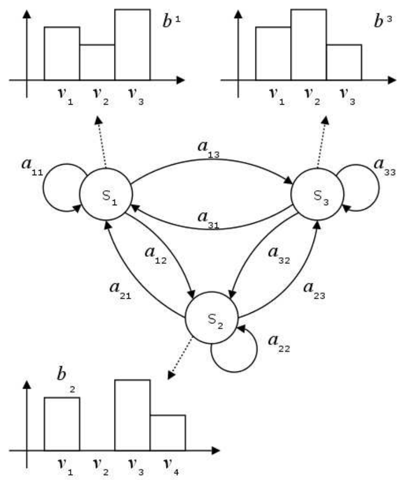

Figure 2: An example of Hidden Markov Model

other given the present state; thus probability of observing symbol v_k at time t in a sequence of symbols does not depend on the symbol observed at time $t-1$, but it depends only on the present state s_i and, implicitly, the previous one $s_{i'}$ through the transition probability $a_{ii'}$ (Ghahramani, 2001).

HMMs could be adopted to implement a predictive models for Bliss symbols if we could assume that a symbol is predictable given the corresponding Bliss symbol category as the hidden state. However, this approach oversimplify the user language model described previously: it does not relate the emission of a symbol with the symbol previously emitted due to the independence assumption in HMMs. To face this issues we have adopted a particular extension of HMM, called AR-HMM (Auto-Regressive Hidden Markov Model) that relate the emitted symbol both to the actual state (as canonical HMM) and to the previous emitted symbol (Figure 3 illustrates the differences between canonical HMM and AR-HMM). In such a way we have a model that keeps into account the previous emission and it is still computationally tractable as described further on.

In order to identify the possible hidden states of an ad-hoc AR-HMM for the Bliss language, symbols have been divided into six categories according to their grammatic role, and, later, each category has been divided into a number of subcategories adopting the semantic networks formalism (Quillian, 1968) to keep into account the semantic of the symbols and the logic connection among two subcategories. This subcategories identification process has been accomplished in collaboration with experts in verbal rehabilitation to obtain subcategories not excessively specific that would have complicated the model without any reason (e.g., we have a substantive subcategory

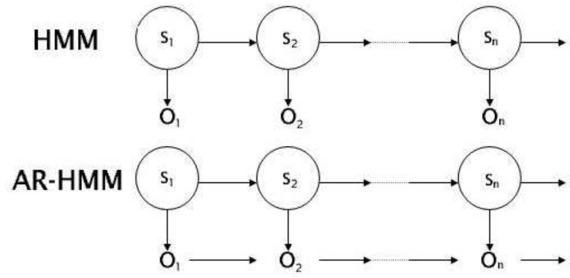

Figure 3: Comparison between HMM and AR-HMM

'food' because it connects the verb subcategory 'feeding', we have not a substantive subcategory 'animal' because it does not connect a specific category). We report such subcategories and the number of symbols assigned to each subcategory (note that a symbol can belong more than one category).

- *Verbs*: people movement (23), objects movement (15), body care (16), description (3), everyday (10), servile (7), emotional (7), feeding (33), other (180).
- *Adverbs*: time (67), exclamatory (12), place (28), quantity (17), holidays (12), color (23), other (20).
- *Adjectives*: opinion (29), character (18), physical description (33), food description (17), quantity (13), feeling (29), other (52).
- *Substantives*: food (141), cloth (38), body (47), everyday (26), people place (110), things place (22), other (600).
- *People*: possession (16), relation (47), job (38), other (51).
- *Punctuation*: question (13), other (36).

4 DISCRETE AUTO-REGRESSIVE HIDDEN MARKOV MODEL

AR-HMMs are commonly used in literature for prediction in continuous systems and they usually describe the emission probability of an symbol/value according to a Gaussian distribution; the emission of Bliss symbol, however, is a discrete event that can be described adopting a multinomial probability distribution. In CABA^{2}L, to overcome this problem, we have implemented a Discrete Auto-Regressive Hidden Markov Model (DAR-HMM) where the emission probability for a symbol is described using a bivariated multinomial distribution. In fact, we introduced the DAR-HMM as a first order extension of a

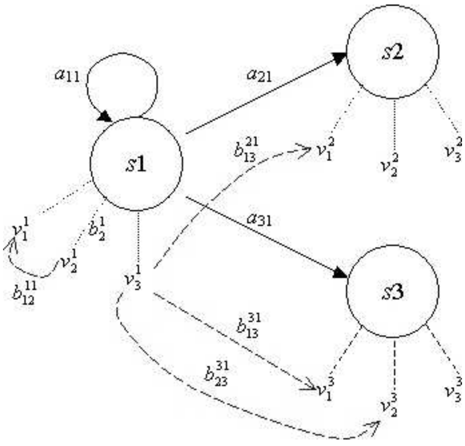

Figure 4: Symbols emission in DAR-HMM; s_i is the state (symbol subcategory), v_j are the observed symbol

classical HMM where the symbol emission probability depends on the present state and the last observed symbol as depicted in Figure 4.

DAR-HMM for symbolic prediction can be described using a parameter vector $\lambda = < \Pi^0, A, B >$, where $\Pi^0[N]$ is the vector of inital subcategory probability $\pi_i(0)$, $A[N][N]$ is the matrix with subcategory transition probabilities $a_{ii'}$, and $B[N][M][M+1]$ is the emission matrix[1] with symbol probabilities $b_{kk'}^{ii'}$ and b_k^i (see Appendix A for details). In CABA^{2}L, this λ vector has been estimated using a dataset of Bliss sentences. To do that, we have adopted a variation of the Baum-Welch algorithm, an iterative algorithm based on the Expectation-Maximization method (Bilmes, 1998; Dempster et al., 1977), adapting this technique to the specific case of DAR-HMM (see Figure 5).

Since the Baum-Welch algorithm is a greedy algorithm that can be trapped in local minima, the initialization estimate of λ parameter vector is a fundamental aspect. In literature a theoretical solution that addresses such issue does not exist; in practice, the adoption of a random or uniform distributed initialization for A and Π^0 has been verified to be adequate. In particular we adopt an uniform distribution as initial estimate for Π^0, and a distribution based on the knowledge about the phenomenon for A. Only arcs connecting subcategories in the semantic model of the language (see Section 3) should have a probability $a_{ii'} \neq 0$. However, we have assigned to the arcs between symbols and states that are not connected in the semantic network a very low probability, not to preclude the training algorithm to eventually discover unforeseen correlations.

The initial estimation for the B matrix is more critical so we have used the *Segmental k-Means* (Juang and Rabiner, 1990; Juang et al., 1986) technique to obtain a more confidential estimate. Such process considers a sub set of sentences composing the dataset, and, for each one, it looks for the best sequence of subcategories using the Viterbi algorithm to upgrades the symbols emission probabilities.

Given the initial values λ^0 for the model parameters, we use a modified Baum-Welch algorithm to estimate, from a real dataset, the model parameters through a sequence of temporary $\overline{\lambda}$ model parameters. As in any learning algorithm, the main issue is avoiding the *overfitting* phenomenon (Caruana et al., 2001), so we would like to stop the training phase according to the generalization error (i.e., the error on new samples) and not just observing the training error (i.e., the error on the training set). To do this, we have used the *K-fold cross-validation technique* (Amari et al., 1995); it consists in dividing the whole set of sentences into K similar subsets to use at each iteration $K-1$ subsets for parameter estimation and the remaining validation set is used to valuate the convergence of model generalization error. In other words, we calculate the errors of the model in predicting the sentences of the validation set it has never seen, and we analyze the validation error function during training iterations of the Baum-Welch algorithm until it reaches its minimum.

In order to terminate the iteration at which the error function reaches its minimum, several practical techniques can be adopted, but none of them assures the achieving the global minimum. We have chosen to adopt a termination criterion based the *generalization loss* method (Prechelt, 1996). Given:

$$Err_{Opt}(t) = \min_{t' \leq t} Err_{Val}(t')$$

the minimum error is obtained at time t; consider

$$GL(t) \triangleq 100\Big(\frac{Err_{Val}(t)}{Err_{Opt}(t)} - 1\Big)$$

which represents the last increment in comparison with the minimum. The training phase is stopped whenever the generalization loss GL becomes bigger than a given threshold τ:

$$GL(t) > \tau.$$

In this approach, the error function could stabilize after a local minimum, without $GL(t)$ rising the threshold. In order to face such issue we have added to

[1] From an implementation point of view matrix B could represent the main issue of this model (i.e., with $N = 30$ subcategories and $M \simeq 2000$ symbols the cells number amount is of the order of 10^8, about 400MBytes). However B can be considered a sparse matrix since from each subcategories only a part of symbols can be emitted, so the cells number is, approximately, lower than 10^4 and ad-hoc data structure such as *heap* or *priority queue* and optimized algorithms can be used to overcame memory occupancy and speed access issues.

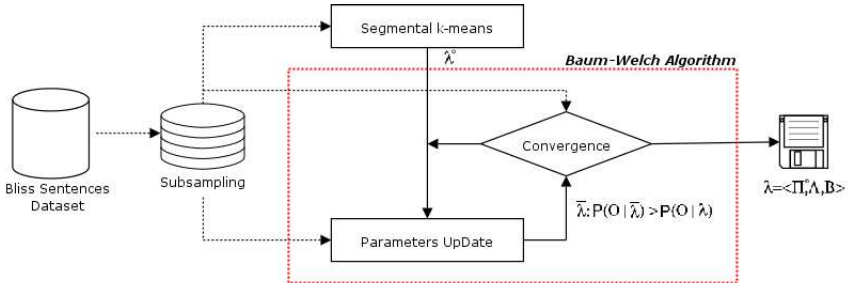

Figure 5: The training process

the stop criterion two condition relating the maximum number of iterations and the minimum improvement during learning.

5 EXPERIMENTAL RESULTS

DAR-HMM has been implemented in CABA^{2}L and, finally, integrated in BLISS2003, a communication software centered on Bliss language. CABA^{2}L receives from BLISS2003 the last selected symbol, calculates the most probable four symbols according to the established requirements, and scans them in an ad-hoc panel in the graphical interface before scanning the full table.

In order to validate DAR-HMM, we are interested in giving an estimated *training error* and *generalization error* in several user scenarios characterized by symbols, symbols subcategories, user residual linguistic capabilities, and user needs; we are also interested in evaluating the time required both for learning and prediction process. To accomplishing this validation, we have strictly collaborated with two Italian clinics for verbal impairments (PoloH and SNPI of Crema[2]) evaluating the prediction performance in different scenarios; in this paper we report just two scenarios as the most significant ones:

1. a dataset of 20 sentences with 4 sub-categories and 7 symbols representing a verbal impaired person unskilled in Bliss utilization or suffering deep mental deficiency;
2. a dataset of 80 sentences with 18 sub-categories and 120 symbols representing a verbal impaired person skilled in Bliss use and without deep mental deficiency.

We have shuffled the sentences of each dataset in order to achieve a homogeneous dataset not affected by time correlation. In addition we have divided each dataset into two parts, respectively 80% of sentences in the first part and 20% of sentences in the second one. We have adopted the first part to training the model computing the training error. We have adopted the second one to evaluate the generalization error.

Table 1: Training error: probability that the requested symbol is in the first four predicted symbols according to the datasets adopted to train the DAR-HMM

Predictions	Scenario 1		Scenario 2	
	Mean	Std. Dev.	Mean	Std. Dev.
1 symbol	0.343	0.055	0.250	0.017
2 symbols	0.563	0.074	0.299	0.028
3 symbols	0.778	0.067	0.319	0.033
4 symbols	0.908	0.056	0.345	0.042
not suggested	0.092	0.056	0.655	0.042

The training error expresses the effectiveness of the learning and it is obtained comparing the suggestion proposed by CABA^{2}L during the composition of sentences it has learnt. To estimate the correct prediction's probability, we have carried out over 800 simulations where we compare the suggested symbols and the one chosen by the user. In Table 1 mean and standard deviation for both the two scenarios are showed, they evidence a training error of about 9.2% for the first scenario and 65.5% for the second one taking into account a number of proposed symbols equals to 4 as suggested by therapist and according to the requirements from Section 3.

The generalization error expresses the effectiveness of the prediction system, and it is obtained comparing the suggestions proposed by CABA^{2}L during the composition of sentences that exhibit the same probability distribution with respect to the sentences it has learnt, but were not presented to the system during the training phase. To estimate the correct prediction's probability, we have carried out over 200 simulations where we compare the suggested symbols and the one chosen by the user. In Table 2 mean and standard deviation for both the two scenarios are showed,

[2]PoloH is a information technology center that support AAC aids adoption. SNPI is the neuropsychiatric adolescent and infancy local service associated to the hospital of Crema, Italy.

they evidence a generalization error of about 11.3% for the first scenario and 64.3% for the second one taking again into account a number of proposed symbols equals to 4 before. The values of mean and standard deviation evaluated in generalization error are very close to the values evaluated in training error, thus DAR-HMM evidences high generalization ability. Although the training and generalization errors are in the second scenario high we are confident to get better result just having a bigger dataset.

Time spent by verbal disables that collaborated with us in order to compose messages using BLISS2003 with respect to the time spent with adoption of a traditional scansion system has been reduced up to 60%. Tests have evidenced that the training phase requires few minutes depending on the size of the dataset and the number of symbols and subcategories, but this does not affect BLISS2003 performance, because it can be run on background. Conversely, these tests have proved that the symbols prediction is immediate ($<$1 second) and can be performed in real time.

6 CONCLUSIONS

In this paper we have analyzed the AAC symbols scansion issues for motor disordered persons establishing requirements according to literature and the experiences of several clinics for verbal disables that have collaborated with us. In particular we described prediction models currently adopted in AAC context and we designed an ad-hoc prediction model (DAR-HMM). We described DAR-HMM peculiarities: its formalism, ad-hoc emission rules, parameters initialization, training processes, stopping criterion, and implementation issues. We have applied DAR-HMM to the case of Bliss language introducing semantic categories for Bliss symbols. In addition, we integrated CABA^{2}L into BLISS2003 an AAC communication software based on Bliss, and experimentally validated it with real data in collaboration with two Italian clinical centers for verbal impaired people proving its effectiveness for reduction of the time spent to compose Bliss messages.

In future the performance of the prediction will be improved refining the prediction model. Moreover we would like to achieve on-line adaptation of the DAR-HMM to the linguistic behavior of the user and to take into account the evolution of the user linguistic capabilities, and to support other AAC languages with respect to Bliss, particularly PCS. Finally we will analyze the learnt semantic/probabilistic model of the linguistic behavior of the user in order to study relationships between disabilities and verbal impairments.

Table 2: Estimated generalization error: probability that the requested symbol is in the first four predicted symbols according to the datasets not adopted to train the DAR-HMM

Predictions	Scenario 1		Scenario 2	
	Mean	Std. Dev.	Mean	Std. Dev.
1 symbol	0.202	0.082	0.185	0.089
2 symbols	0.438	0.146	0.252	0.073
3 symbols	0.666	0.181	0.304	0.070
4 symbols	0.887	0.067	0.357	0.077
not suggested	0.113	0.067	0.643	0.077

REFERENCES

Aliprandi, C., Barsocchi, D., Fanciulli, F., Mancarella, P., Pupillo, D., Raffaelli, R., and Scudellari, C. (2003). AWE, an innovative writing prediction environment. In *Proc. of the Int. Human Computer Conf.*, Crete, Greece.

Amari, S., Finke, M., Muller, K. R., Murata, N., and Yang, H. (1995). Asymptotic statistical theory of overtraining and cross-validation. Technical Report METR 95-06, Dep. of Math. Eng. and Inf., Physics, Uni. of Tokyo, Tokyo.

Bilmes, J. (1998). A gentle tutorial of the EM Algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Markov Models. Technical report, Dep. of Electrical Eng. and Comp. Sci. at Uni. of California, Berkeley.

Bliss, C. K. (1966). *Semantography*. Semantography Blissymbolic Communication, Sidney, Australia.

Bloomberg, K. and Johnson, H. (1990). A statewide demographic survey of people with severe communication impairments. *Augmentative and Alternative Communication*, 6:50–60.

Caruana, R., Lawrence, S., and Giles, C. L. (2001). Overfitting in neural networks: Backpropagation, conjugate gradient, and early stopping. In *Advances in Neural Information Processing Systems*, Denver, Colorado.

Cronk, S. and Schubert, R. (1987). Development of a real time expert system for automatic adaptation of scanning rates. In *Proc. of the Conf. on Rehabilitation Technology, RESNA*, volume 7, pages 109–111, Washington, DC, USA.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society B*, 39:1–38.

Fodor, J. (1983). *The modularity of mind*. MIT Press, Cambridge, USA.

Ghahramani, Z. (2001). An introduction to hidden markov models and bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 1:9–42.

Higginbotham, D. J., Lesher, G. W., and Moulton, B. J. (1998). Techniques for augmenting scanning communication. *Augmentative and Alternative Communication*, 14:81–101.

ISAAC (1983). Int. Soc. for Augmentative and Alternative Communication, Internet site. http://www.isaac-online.org. Last accessed October 1th, 2003.

Juang, B. and Rabiner, L. (1990). The segmental k-means algorithm for estimating parameters of hidden markov models. In *IEEE Trans. on Acoustics Speech and Signal processing*, ASSP-38, pages 1639–1641. IEEE Computer Society Press.

Juang, B., Rabiner, L., and Wilpon, G. (1986). A segmental k-means training procedure for connected word recognition. *AT&T Technical Journal*, 65:21–31.

Koester, H. and Levine, S. (1994). Learning and performance of ablebodied individuals using scanning systems with and without word prediction. *Assistive Technology*, page 42.

Lee, H.-Y., Yeh, C.-K., Wu, C.-M., and Tsuang, M.-F. (2001). Wireless communication for speech impaired subjects via portable augmentative and alternative system. In *Proc. of the Int. Conf. of the IEEE on Eng. in Med. and Bio. Soc.*, volume 4, pages 3777–3779, Washington, DC, USA.

Prechelt, L. (1996). Early stopping-but when? In *Neural Networks: Tricks of the Trade*, pages 55–69.

Quillian, M. (1968). Semantic memory. In *Minsky ed. Semantic Information Processing*. MIT Press, Cambridge.

Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. In *Proc. of the IEEE*, 77, pages 257–286. IEEE Computer Society Press.

Shane, B. (1981). *Augmentative Communication: an introduction*. Blackstone, Toronto, Canada.

Simpson, R. C. and Koester, H. H. (1999). Adaptive one-switch row-column scanning. *IEEE Trans. on Rehabilitation Engineering*, 7:464–473.

Swiffin, A. L., Pickering, J. A., and Newell, A. F. (1987). Adaptive and predictive techniques in a cammunication prosthesis. *Augmentative and Alternative Communication*, 3:181–191.

APPENDIX A

In this appendix, we briefly describe the DAR-HMM according to the formalism adopted by Rabiner in (Rabiner, 1989) to specify classical hidden Markov models:

- $S \triangleq \{s_i\}$, subcategories set with $N = |S|$;
- $V \triangleq \{v_j\}$, predictable symbols set with $M = |V|$;
- $V^{(i)} = \{v_k^{(i)}\}$, set of symbols predictable in subcategory i with $M^{(i)} = |V^{(i)}|$ and $V = \bigcup_i V^{(i)}$;
- $O(t) \in V$, observed symbol at time t;
- $Q(t) \in S$, state at time t;
- $\pi_i(t) = P(Q(t) = s_i)$, probability that s_i is the actual subcategory at time t;
- $a_{ii'} = P(Q(t+1) = s_i|Q(t) = s_{i'})$, transition probability from $s_{i'}$ to s_i;
- $b_k^i = P(O(0) = v_k^{(i)}|Q(0) = s_i)$, probability of observing $v_k^{(i)}$ from subcategory s_i at $t = 0$;
- $b_{kk'}^{ii'} = P(O(t) = v_k^{(i)}|Q(t) = s_i, O(t-1) = v_{k'}^{(i')})$, probability of observing $v_k^{(i)}$ from the subcategory s_i having just observed $v_{k'}^{(i')}$.

DAR-HMM for symbolic prediction can thus be described using a parameter vector $\lambda = <\Pi^0, A, B>$, where $\Pi^0[N]$ is the vector of inital subcategory probability $\pi_i(0)$, $A[N][N]$ is the matrix with subcategory transition probabilities $a_{ii'}$, and $B[N][M][M+1]$ is the emission matrix with symbol probabilities $b_{kk'}^{ii'}$ and b_k^i. Given λ the vector of parameters describing a specific language behavior model, we can predict the first observed symbol as the most probable one at time $t = 0$:

$$\begin{aligned}\hat{O}(0) &= \arg\max_{v_k^{(i)}} \left(P(O(0) = v_k^{(i)}|\lambda)\right) \\ &= \arg\max_{v_k^{(i)}} \left(P\left(O\left(0\right)|Q\left(0\right),\lambda\right)P(Q(0))\right) \\ &= \arg\max_{v_k^{(i)}} \left(b_k^i \cdot \pi_i(0)\right).\end{aligned}$$

Then mimicking the DAR-HMM generative model, to predict the t^{th} symbol of a sentence we want to maximize the symbol probability in the present (hidden) state given the last observed symbol:

$$P\left(O(t) = v_k^{(i)}, Q(t) = s_i|O(t-1) = v_{k'}^{(i')}, \lambda\right).$$

Recalling that we can compute the probability of the current (hidden) state as:

$$\begin{aligned}P(Q(t)) &= \sum^N P\left(Q(t)|Q(t-1)\right)P\left(Q(t-1)\right) = \\ &= \sum_{i'=1}^N \pi_{i'}(t-1)a_{ii'} = \pi_i(t),\end{aligned}$$

we obtain a recursive form for symbol prediction at time t:

$$\hat{O}(t) = \arg\max_{v_k^{(i)}} \left(b_{kk'}^{ii'} \cdot \sum_{i'=1}^N \pi_{i'}(t-1)a_{ii'}\right).$$