Multimed Tools Appl
DOI 10.1007/s11042-012-1073-5

Making assistive reading tools user friendly: a new platform for Greek dyslexic students empowered by automatic speech recognition

Theologos Athanaselis · Stelios Bakamidis · Ioannis Dologlou · Evmorfia N. Argyriou · Antonis Symvonis

Abstract This work presents our effort to incorporate a state of the art speech recognition engine into a new platform for assistive reading for improving reading ability of Greek dyslexic students. This platform was developed in the framework of the Agent-DYSL, IST project, and facilitates dyslexic children in learning to read fluently. Unlike previously presented approaches, the aim of the system is not only to enable access to the reading materials within an inclusive learning system but to promote the development of reading skills by adjusting and adapting in the light of feedback to the system. The idea is to improve speech recognition performance so that gradually increase the reading capabilities of the user, gradually diminish the assistance provided, till he is able to read as a non-dyslexic reader. The evaluation results show that both learners' reading pace and learners' reading accuracy were increased.

Keywords Assistive reading tool · Greek dyslexic students · Speech recognition engine · Vocal user interface

T. Athanaselis (✉) · S. Bakamidis · I. Dologlou
Voice and Sound Technology, Institute for Language and Speech Processing (ILSP)/R.C. "Athena",
Artemidos 6 & Epidavrou, 151 25 Maroussi, Athens, Greece
e-mail: tathana@ilsp.gr

S. Bakamidis
e-mail: bakam@ilsp.gr

I. Dologlou
e-mail: ydol@ilsp.gr

E. N. Argyriou · A. Symvonis
Department of Mathematics, School of Applied Mathematical and Physical Sciences,
National Technical University of Athens,
Iroon Polytexneiou 9–Zografou, 15780 Athens, Greece

E. N. Argyriou
e-mail: fargyriou@math.ntua.gr

A. Symvonis
e-mail: symvonis@math.ntua.gr

Published online: 12 May 2012

1 Introduction

Dyslexia is a specific learning difficulty that primarily affects reading and spelling. It is lifelong, although its effects can be minimised by targeted literacy intervention, technological support and adaptations to ways of working and learning. Dyslexia is characterized by difficulties in processing word-sounds and its effects may be seen in spoken language as well as written language.

Dyslexia has still not achieved full recognition in all EU countries, with provision in schools being highly variable [24]. It is estimated that about 3 million children in the E.U. and 4.5 million children in the whole of Europe are severely dyslexic and, therefore, low achievers in reading and learning. Furthermore about 30 million adults in the E.U. (out of 460) and about 55 million people in Europe (out of 730) have some problems in reading due to dyslexia and other Specific Learning Difficulties. The loss of opportunity–in both human and economic terms–resulting from dyslexia is considerable [16, 26].

The intent of the paper is to present the overall Agent-DYSL system [1] and to focus on its automatic speech recognition module. The development of the speech recognition module is driven by the modelling of the Dyslexia for the Greek language using several types of *reading errors*. Based on the reading errors the vocabulary of the speech recognition module is developed. Agent-DYSL proposes a new system which utilizes voice recognition and image processing to monitor the engagement of the learner and to aid in the identification of individual learning needs. The goal of the system is not only to enable access to the reading materials within an inclusive learning system but to promote the development of reading skills by adjusting and adapting in the light of feedback to the system. In this way it aims to be both enabling and instructional, developing independent learning skills and supporting inclusive learning across a range of environments [18].

The remainder of this paper is organized as follows: Section 2 describes briefly the literature review. Section 3 includes the architecture of the entire system and a description of user interface. Section 4 shows how a state of the art speech recognition engine can be incorporated into the Agent-DYSL system. Section 5 describes the experimentation and Section 6 presents the Speech recognition performance and the Learner's Improvement using Agent-DYSL as well. Finally, Section 6 discusses the system's performance, following by Section 7 with concluding remarks and future plans.

2 Literature review

2.1 Assistive reading tools

The benefits of using assistive computer software to help dyslexics with reading difficulties have been recognized [7, 8] and there are several reading assistance commercial systems like AllWrite [2], Kurzweil 3000 [13], LexiaReading [14], ClaroRead [5] RapidReaders [19], ReadOutloud Solo [20], ReadOn Company [21], RocketReader [22], SpeakOut [25] by Sonant specifically targeted towards dyslexia, now available. State of the art commercial software applications hardly have any form of user adaptivity beyond simple preferences. Recent research approaches like [3, 12] concentrate particularly on dyslexia and make use of speech recognition and eye tracking to adapt to readers' progress. However, all of these approaches do not allow for deep adaptation by, e.g., considering user-specific error types. This way, pedagogical intervention is very limited.

2.2 Speech recognition for dyslexia

In the literature many systems for CALL (Computer-assisted language learning) have been described that incorporate speech technology. Pioneering research by MIT and CMU as well as more recent work by IBM Watch-me-Read have demonstrated that speech recognition can play an effective role in systems designed to improve children's reading abilities [32]. Research on reading tutors started long ago e.g. LISTEN system by Mostow et al. [17] and the STAR system by Russel et al. [23] e.g. within the Flemish SPACE project [6]. Interesting work has also been carried out in the Foundations to Literacy reading program [30], in which virtual tutor- a 3D computer model- interacts with children in multimodal learning tasks to teach them to read. A key component of this programs is the Interactive Book, which combines real time speech recognition, facial animation and natural language understanding capabilities to teach children to read and comprehend text.

None of these programs have taken into consideration either the learner's current state and performance or save it and restore it the next time the learner uses the software. As a result, these programs cannot provide personalized assistance. This can be traced back to the problem that they only rely on a small set of technologies (usually highlighting and speech output). But only a combination of different technologies in an intelligent way will provide software that can be adaptive to its learners' needs, which includes speech recognition, and image recognition to get information about the learner, which is a prerequisite for a personalised learning environment, which takes into account the individual needs of each dyslexic learner.

Despite the fact that several assistive reading software packages are available, there are some prominent features that are absent from these tools. The key innovation of the Agent-DYSL reading assistance is the combination of (1) a deep knowledge about the user and the current context or situation, (2) context-aware pedagogical knowledge about how to respond to a particular situation and (3) appropriate adaptation possibilities in the reading software user interface. Furthermore, the system allows for teachers and experts to gain new insights into the effectiveness of certain pedagogical strategies.

3 Agent-DYSL system architecture

Agent-DYSL provides an adaptive reading assistance system for dyslexic students which allows them to read arbitrary text documents. The system presents the text in an augmented way by using techniques such as text highlighting, segmenting words into its syllables, putting emphasis on certain characters, or pre-emptively reading words aloud by the usage of text-to-speech techniques. The architecture of the proposed system consists of four main components (see Fig. 1):

a) Recording and Analysis Component (consisting of image and speech recognition),
b) Knowledge Infrastructure,
c) Profiling component, and
d) Content Presentation component.

The components are used to assess the learner's profile that will personalize the reading environment. In this mode the learner is asked to read aloud a text which is presented in the software window while both microphone and camera are activated. The data from the microphone and the camera are analysed using speech recognition and face analysis (image recognition) respectively. Speech recognition and face analysis algorithms are parts of the *Recording and Analysis component*. Image recognition process requires a number of

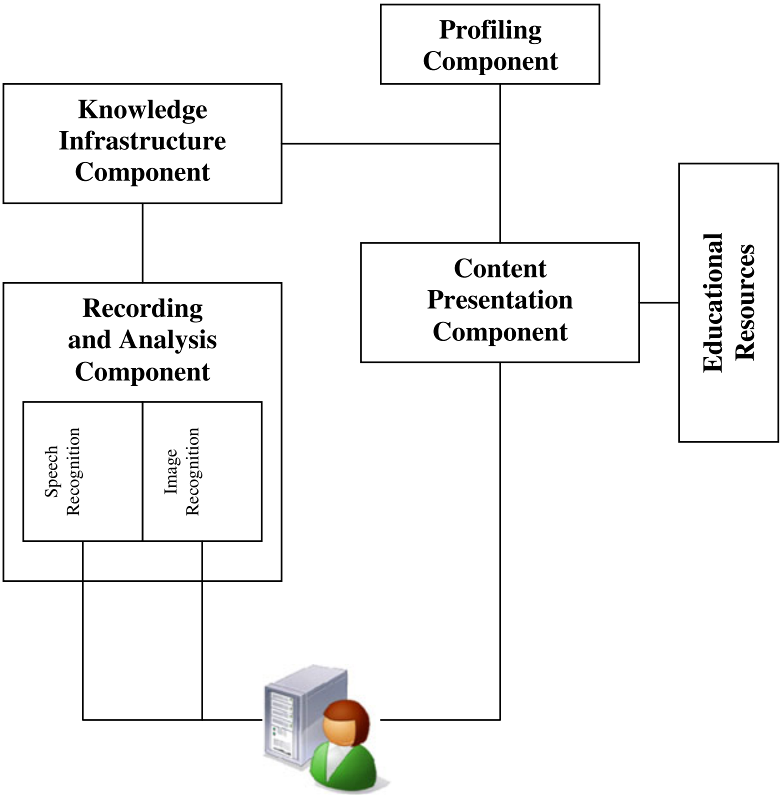

Fig. 1 System architecture

pre-processing steps, which attempt to detect or track the face, to locate characteristic facial regions such as eyes, mouth and nose on it, to extract and follow the movement of facial features, such as characteristic points in these regions.

The performance of these algorithms is further evaluated by the *Knowledge Infrastructure* component. The *Knowledge Infrastructure* component is considered as the manager of the proposed system and it is responsible to provide the system with the learner's error profile. With the help of state-of-the-art ontology-based techniques like OWL and rule languages like SWRL, expert pedagogical knowledge can be encoded in a descriptive way. Part of this expert knowledge takes the form of rules that take contextual variables from the user model as an input and yield adaptation actions as an output (*Knowledge Infrastructure*). This encoded knowledge will certainly only represent a subset of teachers' experience, but it still yields a substantial improvement compared to traditional reading software. It will be stored in a central server, along with the knowledge derived from users' profiles. User will have to log on this server, in order to take advantage of all the encoded knowledge, according to their specific needs. The learner's error profile is stored into the *Profiling* component.

The *Profiling* component includes not only individual error profiles but also individual preferences which may have been defined by the learner. The Profiling component uses the Knowledge Infrastructure both for storage and retrieval. At the end, the learner's current profile is stored into the Profiling component.

The *Content Presentation* component receives a file in PDF format as input and after proper analysis, the text and the images from it are extracted and an XML file is produced. With the help of the Knowledge Infrastructure, this text is analyzed with respect to words that are likely to cause problems to the learner. For these words, appropriate changes to the presentation are determined (e.g., augmented with specific highlighting colour and speed, word and line spacing) [27].

In addition, the proposed system provides several multimedia benefits such as text highlighting, line spacing, adjusting fonts, segmenting words into its syllables, emphasis on certain characters, or pre-emptively reading words aloud by the usage of text-to-speech techniques. The learners' profiles and preferences data are retrieved either from the Agent-DYSL server if learner's PC has an internet access or from a USB stick (in which profiles and preferences were stored from the learner in a previous usage of the application). Also provides a teacher tool in order to change the learner account settings such as password, learner preferences and learner profile, can write down his/her assessment for the learner, see a description of the current learner profile or print a report with recommendations associated with the profile of the learner. The system can also suggest individual learning resources for further training, which is useful for a teacher of a dyslexic child.

4 Agent-DYSL automatic speech recognition module

Automatic Speech Recognition (ASR) main task is to record (in text form) the user's voice, while he reads aloud a given text. Specifically, the ASR performs the task of identifying word mispronunciations and reading errors. The ASR module contributes to the following objectives of Agent_DYSL:

- Automatic user progress monitoring: The ASR will be able to recognise the user's reading and detect discrepancies from the original text, enabling user evaluation and progress monitoring.
- Automatic user modelling: By detecting and recording the user's reading errors, the ASR will enable the Agent-DYSL intelligent reading system to predict the reading errors the user is likely to make, but also take into account the user's improvements in updating the user profile.

4.1 Related work

The ASR system vocabulary should be composed of a set of highly frequent words to increase the recognition rate. The number of words in the vocabulary is subject to several constraints like the recognition speed. A large vocabulary would slow down speech recognition causing problems to the entire system since the ASR output interacts online with other modules of the Agent-DYSL system. However, the number of words is enough to cover the various dyslectic phenomena and at the same time to provide a reliable evaluation of the user's reading performance. Related to previous work, Hulslander [11] (Olson Reading Lab.) derived a classification of reading errors that was used to annotate a corpus for training a speech recogniser used in Project LISTEN [9]. In project LISTEN, the recogniser was used to monitor a child reading aloud so that the system could judge when a child was making mistakes and when to interrupt him/her. Their classification scheme identifies a large number of types of errors grouped under the headings of substitutions, insertions, omissions, fluency errors, repetitions and self-corrections. In [28, 29] reading error data was different to Olson

Lab's. Their classification was similar to van Hasselt's [10], whereas her study measured numbers of errors, but they also measured times of errors.

The approach we followed in the selection of our vocabulary is rather unique compared to previous related work. Children with dyslexia make specific reading errors which have been classified into several error types. The error types depend on the structure and complexity of each word. For example, there are error types concerning multi-syllable words, composite words, words containing special combination of consonants, etc. Based on the error types, a profile is created for each child which (the profile) reflects the degree of "difficulty" each error type causes to the child with dyslexia. The need to evaluate the degree of difficulty caused by each error type dictated the creation of a list of relevant words to each error type (these words were referred to as "sensitive" words). Given that the users of the system were third and fourth grade primary school children, the sensitive words were chosen from the relevant reading text-books used in public schools. Then, these sensitive words were used in simple sentences appropriate so that each sentence contained one sensitive word and was only relevant to one error type. These sentences were presented to the children during the evaluation/creation of their profile. Based on the above description, the vocabulary of our system consisted of the sensitive words and the extra words used to make up sentences. The size of the vocabulary was proportional to the number of error types. For each error type relevant to the appearance of a specific consonant, 10 sensitive words were selected. For the error types relevant to complex words and composite words, a larger number of sensitive words was selected in order to account for the different pronouns used as the first component in the composite words. All together, the resulting list of sensitive words consisted of 258 words. In addition, 1,063 words were used in making up sentences, resulting to a vocabulary of 1,321 words.

4.2 Speaker adaptation module

It is known from the literature that some of the spectral properties of speech are due to the Vocal tract Length (VTL). According to the linear acoustic theory of speech production, the formants in speech are VTL-dependent. Consequently, one has to warp the average spectrum represented by the HMM models to match it with the input spectrum.

This is achieved by computing the third formant F3 of the current speaker and comparing it against the average values of the 3rd formant of the training speakers. VTL then compensates for the existing differences between these values. Warping techniques have been implemented to adjust the VTL of the speaker to the average VTL of the training speakers.

VTL normalization has been used for efficient speaker adaptation. The method is based on the automatic normalization of the speaker VTL to the average VTL of the training speakers.

The VTL is related to the i-th average formant with (mean) frequency Fi as

$$VTL \approx \frac{(2i-1)c}{4F_i}$$

Where c is the velocity of the sound.

From the above equation we can assume that:

$$\frac{VTL_{average}}{VTL_{input}} \approx \frac{F_{i,input}}{F_{i,average}}$$

Therefore, normalization of the VTL can be achieved by a frequency warping procedure.

Speaker adaptation is achieved by progressively adjusting the VTL of the speaker to the average VTL of the speakers of the training set.

4.3 Description of automatic speech recognition (ASR) module

The present module provides the ability to process the speech signal and to perform feature extraction, by converting each speech frame into a set of cepstral coefficients. It then concentrates on acoustic modelling, by allowing acoustic phoneme models to provide features probability estimates, with a given sequence of words (see Fig. 2).

4.3.1 System overview

The input speech waveform (typically sampled at 16 kHz) is first analysed into a sequence of acoustic feature vectors such as the Mel Frequency Cepstral Coefficients (MFCCs). A popular choice is the first 12 cepstral coefficients and an overall energy feature together with first and second time derivatives of these features, giving a 39-element vector.

Once the input speech has been analysed into a sequence of feature vectors, the recognition task is to find the most probable word sequence Ŵ given the observed vector sequence Y. Revisiting Bayes' theorem, but applying it to the task of finding a word sequence, the most probable sequence can be derived from the probability P(W|Y) of any of one sequence W as follows :

$$\widehat{W} = \arg\max_{W} P(W/Y) = \arg\max_{W} \frac{P(W)P(Y/W)}{P(Y)} \tag{1}$$

Equation (1) states that the most likely word sequence is the one which maximizes the product of P(Y|W) and P(W). The first term denotes the probability of observing vector sequence Y given the word sequence W, and is determined by an acoustic model. The second term represents the probability of observing word sequence W independently from the acoustic signal, and is determined by a language model. For all but the simplest applications, the language-model probability is also a major factor in obtaining good performance: restrictions imposed by the language model can greatly reduce the number

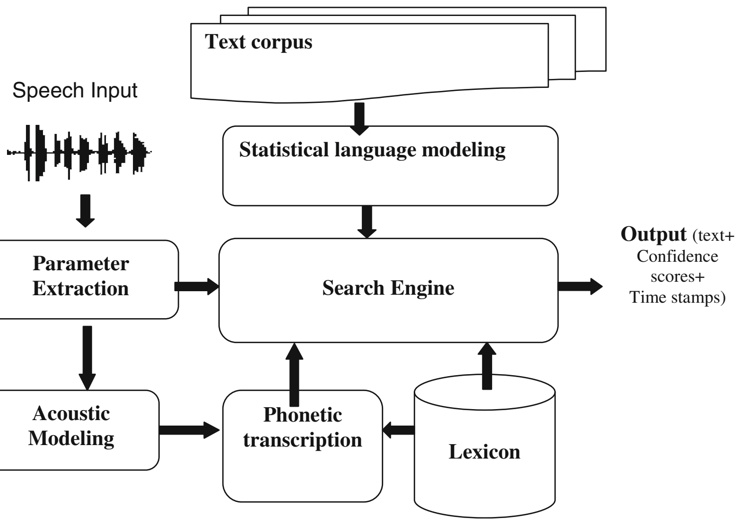

Fig. 2 The speech recognition module

of different alternatives to be distinguished by the acoustic model. As with the acoustic model, the language model is usually a statistical model that is automatically trained on data. In the case of the language model, these data usually take the form of text material chosen to be representative of the recognition task. Figure 3 illustrates the main components of the speech recognition module [4, 31].

4.3.2 Parameter extraction

The prime function of the parameter extraction module is to divide the input speech into blocks; then for each block a smoothed spectral estimate is derived. The spacing between blocks is typically 10 msecs and blocks are normals ly overlapped to render a longer analysis window, typically 25 msecs. It is quite usual to apply a tapered window function (e.g. Hamming) to each block. The required spectral estimates may be computed via Linear Prediction or Fourier analysis. In fact, there are a number of additional transformations that can be applied for the generation of Mel-Frequency Cepstral Coefficients (MFCCs), based on the Mel-scale, designed to approximate the frequency resolution of the human ear, which is linear up to 1,000 Hz and logarithmic after that point. After computing the spectral estimates, Discrete Cosine Transform (DCT) to the log filter-bank coefficients is performed. This process has the effect of compressing the spectral

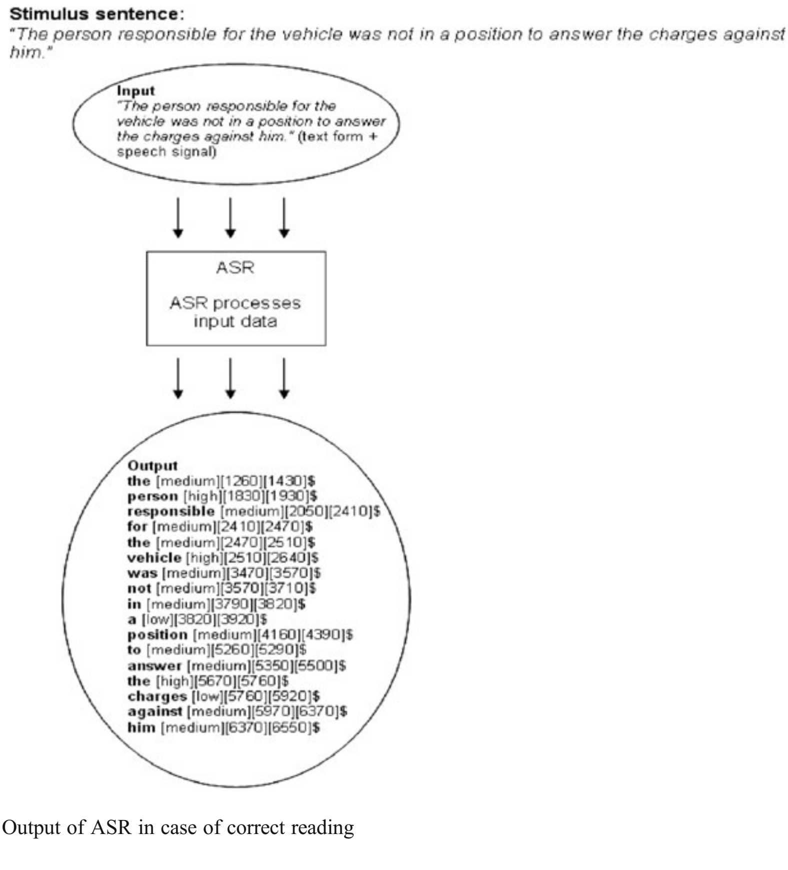

Fig. 3 Output of ASR in case of correct reading

information into the lower order coefficients and it also de-correlates them so that the subsequent statistical modelling can use diagonal covariance matrices. Cepstral coefficients can also be derived from LP coefficients where they achieve a similar de-correlating effect. Flexible pre-processing of the specific recognition engine results to the possibility of a user configurable pipeline of simple processing blocks and available block processing for all subsequent common operations (FFT, mean-normalisation, Mel-filterbanks, time-derivatives) [31].

4.3.3 Acoustic modelling

The purpose of the acoustic models is to provide a method of calculating the likelihood of any vector sequence given a word w. Word sequences are decomposed into basic sounds called phonemes. Each individual phoneme is represented by an HMM. HMM phoneme models typically have three emitting states and left-to-right topology. The Greek ASR uses 32 phonemes to describe the pronunciation of all words [31]. The LOGOTYPOGRAFIA [15] speech database (completed by ILSP in 2001) was used for training the corresponding HMMs. These acoustic models were adapted to child voice specifications with the use of speech samples taken from 22 child speakers aged 9+/−1 years.

4.3.4 Statistical language modelling

The purpose of the language model is to provide a mechanism for estimating the probability of some word wk in an utterance given the preceding words W_1^{k-1} . An effective way of doing this is to use N-grams in which it is assumed that wk depends only on the preceding N-1 words, that is

$$P\left(w_k/W_1^{k-1}\right) = P\left(w_k/W_{k-n+1}^{k-1}\right) \tag{2}$$

N-grams simultaneously encode syntax, semantics and pragmatics and they concentrate on local dependencies. This makes them very effective for languages like English where word order is important and the strongest contextual effects tend to come from near neighbours. Furthermore, N-gram probability distributions can be computed directly from text data and hence there is no requirement to have explicit linguistic rules such as a formal grammar of the language. In principle, N-grams can be estimated from simple frequency counts and stored in a look-up table [4, 31].

4.3.5 System vocabulary

The system vocabulary for Greek includes 1,321 words, both error-sensitive words and sentence words (filler items). The words were again drawn from school text books addressing the ages that Agent-DYSL targets (i.e. 9–11 years). The sensitive words were selected based on their sensitivity to the following error types:

- Phoneme visual confusion
- Errors in polysyllabic words
- Errors in composite words
- Errors in complex structure words
- Stress errors

Sentences A number of sentences were created to include the sensitive words. Three sentences were created for each sensitive word, while each sentence contained only one sensitive word. The sentences are of age-appropriate complexity and length (between 4 and 6 words). A sample list of the sentences used for training ASR language model is shown above in Tables 1 and 2.

Error variants A list of 228 alternative words was also introduced in the system as possible error variants of the sensitive words. Like in English, the error variants were selected through trial testing, where children with dyslexia were asked to read the system sentences while their errors were recorded. Any mispronunciations of several sensitive words were included in the error variants list, which is available in Table 3.

4.3.6 Search engine

The basic recognition problem is to find the most probable sequence of words given the observed acoustic signal (based on the Bayes' rule for decomposition). In our system, we use the breadth-first approach and specifically, beam search and Viterbi decoding (it exploits Bellman's optimality principle).

The dynamic performance in this search engine accomplishes a system capable of exploiting complex language models and HMM phone models depending on both the previous and succeeding acoustic context, such as co-articulation. Moreover, it can do this in a single pass, in contrast to most other Viterbi-systems that use multiple passes.

4.3.7 Output of speech recognition module using confidence measures

The speech recognition system will not only provide a single transcription of a spoken utterance but also a probability for each word. This probability shows the confidence score of each word comparing to the language model.

Confidence scores are provided by most speech recognisers to give a measure of how confident the result is in the hypothesised recognitions. Confidence scores are given in the range 0–1. The indication of confidence scores is important, since errors occur frequently in speech recognition and can make applications such as spoken dialogue systems too cumbersome to use. A confidence score can help in deciding if a recognition result is inaccurate, thus taking the appropriate action.

5 Interacting with agent-DYSL

The speech recogniser will be able to record the speaker's utterances in text form and detect reading errors. This task will be performed in the way described below.

Step (1) Teacher or parent adjusts recording volume using Windows tools and specialised, high quality microphone

Table 1 The vocabulary of the Greek speech recogniser

Sensitive words	258
Sentence words	1063
Total:	1321

Table 2 Sensitive words and sentences

Sensitive word	Sentence
θάρρος	Ο Κώστας έχει θάρρος.
θάρρος	Το θάρρος σώζει ζωές.
θάρρος	Εάν έχεις θάρρος, θα τα καταφέρεις.
θεόρατα	Αυτά τα σπίτια είναι θεόρατα.
θεόρατα	Κοίτα τα θεόρατα βουνά.
θεόρατα	Αγόρασα κάτι θεόρατα παπούτσια.
θηλυκό	Το θηλυκό σκυλάκι γέννησε.
θηλυκό	Η κόρη μου είναι ζωηρό θηλυκό.
θηλυκό	Δεν έχω θηλυκό ζώο στο κοπάδι μου.
θυμώνω	Όταν θυμώνω κλαίω.
θυμώνω	Δε θυμώνω με τα μικρά παιδιά.
θυμώνω	Όταν θυμώνω φωνάζω δυνατά.

Step (2) Child places microphone at a convenient position, not more than 10 cm from his/her mouth

Step (3) Child presses a button on the computer keyboard and begins reading the sentence that appears on screen (the room must be relatively quiet for the recognition to succeed)

Step (4) The speech signal is recorded and processed by the ASR along with the stimulus sentence in text form—the speech signal and the stimulus sentence in text form comprise the input to the ASR system.

Step (5) After processing the input, the ASR will be able to produce output in the following form:

<word that should have been uttered><[confidence level (low, medium, high)]><[start point (ms)]><[end point (ms)]><[error variant (if any)]><$>

Where $ is just an end marker.

An error type appears only if there exists as an alternative to the word that should have been uttered and delivers better confidence level.

Table 3 Vocabulary integrated in the Greek speech recogniser

Sensitive word	Error form 1	Error form 2	Error form 3
θέαμα	θέμα		
θάρρος	βά		
θεόρατα	θέματα	θεά	θεάρα
οικονομία	οικογέν	οικογένεια	
ανέμελα	ενημε	ανάμεσα	ανέμεσ
υποδειγματικός	υποδεγμετικός		
υποτάχθηκαν	υποκατάχθηκαν	υποτάχθηκες	υποτάχθηκε
αποφάσισε	αποφασίσει	αποφασίζει	
απογοητεύω	απογοητεύομαι	απογοητεύθηκα	απογοητεύεται
αποδυτήρια	αποδυτήριο	αποδυμοτήρια	
αποκάλυψη	αποτάμλυψη	αποκαλύψεις	

To illustrate, let us assume that the user sees on screen and utters «The person responsible for the vehicle was not in a position to answer the charges against him». If the speech recognizer decides that a word was mispronounced and if there are alternatives for this word incorporated in system dictionary, the speech recognizer will suggest which alternative (if any) matches best the mispronounced word. This is illustrated in Fig. 3 (correct reading) and Fig. 4 (incorrect reading) where the word "charge" was wrongly read as "charges".

6 Detecting reading errors for Greek

The evaluation of the Agent-DYSL system started in Greece in December 2008, when the first prototype of the system was made available to teachers for experimentation. A full version of the system was released in March 2009. The system was installed in 6 public primary schools (used with 23 children in total) and in 1 private speech pathology center (used by 10 children). The children from the public primary schools that participated in the program follow inclusion classes that operate in parallel with regular class, according to the mainstreaming policy of the Greek system of education. The main target group included primary school-age children at the age of 9 (+/−1 year). This corresponds to 3rd–5th grade class of the Greek primary school (3rd grade class: 10 children, 4th grade class: 10 children, 5th grade class 13 children)

The schools which participated in the program were selected with the help of the Greek Ministry of Education. In June 2008, teachers involved in special education from several schools on the Athens' region were encouraged to participate in the evaluation of the Agent-

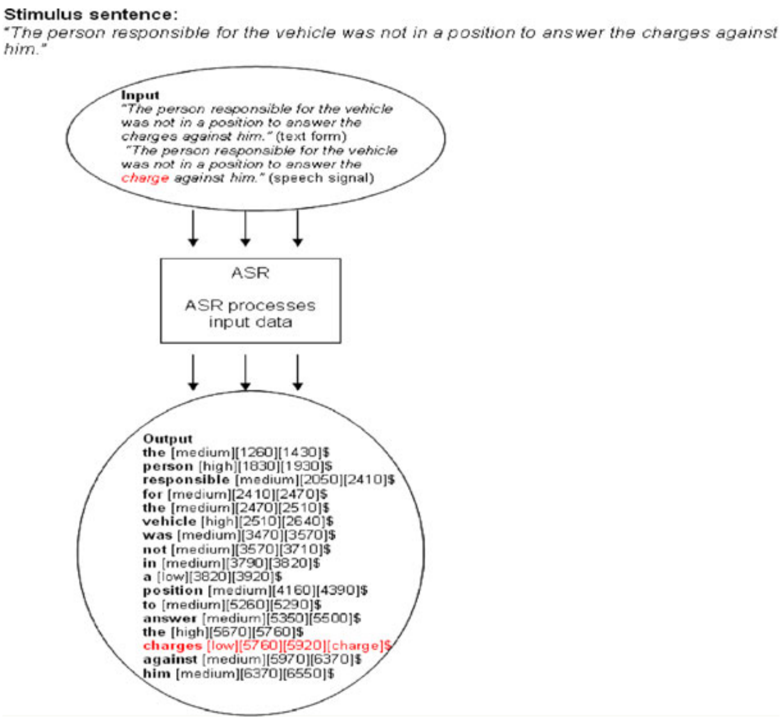

Fig. 4 Output of ASR in case of incorrect reading

DYSL system. Teachers, who expressed interest, were invited at a seminar on the Agent-DYSL system. The seminar took place in the National Technical University of Athens, in June 2008. At the end of that seminar, teachers were asked to report the number of children in each school that could participate in the program (number of children with dyslexia certified by a Diagnostic Evaluation and Support Centre (KDAY)) and whether or not their school had the necessary equipment to run the Agent-DYSL system. In December 2008 after having chosen the schools that fulfilled the requirements for running the system and after having the permission of all headmasters of the participated schools, a prototype version of the system was released to the teachers and their training was started over. Teachers that participated in the program were trained on how:

- To add a new user.
- To create a profile manually.
- Load a document.
- Adjust the settings (highlighting speed for each user, colors, font-size, etc.) for each user.
- Update the profile manually or by doing a re-evaluation automatically.
- Create their own documents and convert them to pdf in order to open them with the Agent-DYSL.

People responsible for the training were in contact with the teachers and visited the schools very often in order to explain and help the teachers that faced difficulties in using the system. Teachers that participated in the program had to keep a log regarding the hours that used the system for each child, and to fill a questionnaire regarding the performance of each child before and after the use of the system and the usability of the system, in general. In Greece, teachers were asked to fill the questionnaire twice. The first time was in April 2009, after using the program for almost 5 months (including some pauses due to technical problems, broken PCs, etc.) and the second time was in June 2009 when the school year had finished.

6.1 Evaluation data on the learners' performance

The main goal of this section is to present the results of the Agent-DYSL system evaluation on the learners' performance. More precisely, we study (before and after the use of the Agent-DYSL system) the reading pace and accuracy of the readers, the intrinsic motivation, the self esteem and the relevance of the software. The results are summarized the results in the Tables 4, 5, 6, 7 and 8. According to teacher's judgment, learner's reading pace after the use of the Agent-DYSL system was increased by 0.52 units (in a scale of 1 [very poor] to 5 [excellent]) after the first evaluation and by 0.56 after the second evaluation. Furthermore, learner's reading accuracy has also been increased by 0.56 after the first evaluation and by 0.61 after the second one.

Table 4 Average data on children's performance after 1st evaluation

	Difference value	Number of answers	Average results in difference
Reading pace	13	25	0.52
Accuracy	14	25	0.56
Motivation	18	25	0.72
Self-esteem	14	25	0.56
Relevance	8	25	0.32

Table 5 Average data on children's performance after 2nd evaluation

	Difference value	Number of answers	Average results in difference
Reading pace	13	22	0.59
Accuracy	13	21	0.61
Motivation	22	21	1.05
Self-esteem	21	21	1.00
Relevance	5	21	0.23

The results after the use of the system are more positive when we study the motivation of children before and after the use of the system. The results are 0.72 and 1.05 units for the first and the second evaluation, respectively. Similar were the results concerning the self-esteem of children; the results show an improvement of 0.56 units after the first evaluation and 1.00 after the second evaluation. As far as the relevance of the software is concerned, the results point out an improvement of 0.32 units after the first evaluation and 0.23 after the second evaluation.

The average use of the system in public schools was 638.56 min or 11.04 h for each child, whereas in private center was 51.3 min.

The Fig. 5 presents the impact of Agent-DYSL system in students' reading performance according to the reading pace, accuracy, motivation self-esteem, and relevance. The results show, the significant improvement using such a system that takes students' needs into consideration.

6.2 Speech recognition performance

Evaluation data were collected during the Agent-DYSL testing for Greek. Quantitative analysis showed that the Speech recognition failure rate drops to 10 % in words longer than 3 syllables while it rises to 36 % in one or two-syllable words. The following table illustrates the results.

In order to reduce the number of failures, a new algorithm was derived that combines short words with longer ones to generate new word constructions, which are more accurately recognized. The correct word in the recognized construction is then extracted by special processing that exploits the prior knowledge of what has been uttered. Using the new enhanced algorithm the results were considerably improved and the failure rate dropped from 36 % to 21 % as shown below.

The table above summarises the speech recognition results per Grade Class.

Table 6 Speech recognition performance in short/normal words

Typical words	Number of trials	Failures	%
ζύμη (zimi) ρίζα (riza) ταψί (tapsi) δένω (ðeno)	71	26	36
τραγούδι (traγuðι) κάθομαι (kaθome) αληθινά (aliθina)	90	9	10

Table 7 Speech recognition performance (WER%) in short words after fine tuning

Typical words	Number of trials	Failures	%
ζύμη (zimi)	71	15	21
ρίζα (riza)			
ταψί (tapsi)			
δένω (ðeno)			
τραγούδι (traγuðι)	90	9	10
κάθομαι (kaθome)			
αληθινά (aliθina)			

7 Conclusions

The main goal of this work is to present the Greek version of Agent-DYSL system and its evaluation results regarding students' performance. In the 2nd evaluation, according to teachers' judgment, learners' reading pace after the use of the Agent-DYSL system was increased by 0.59 units and the learners' reading accuracy has also increased by 0.61. The experimental results showed that the Speech recognition failure rate drops to 10 % in words longer than 3 syllables while it rises to 36 % in one or two-syllable words. On the other hand using the new enhanced algorithm the results were considerably improved and the failure rate dropped from 36 % to 21 % as shown below

Another important feature of the Agent-DYSL system is the extensibility. It should be noted that the development of a new Agent-DYSL system target's language is not a simple localization process. It involves expert data related to each specific language (error type/ presentation rules/data used for the re-evaluation of the user profile), and specific voice data required for the training of the speech recognizer. The development of a new language is based on the same development plan used for Greek language, which consisted of the following steps:

- Identification of error types
- Creation of a list of sensitive words, each relevant to a particular error type
- Creation of a list of alternative "words", each related to a sensitive word.
- Creation of a list of sentences, involving the sensitive word, to be used for the user profile re-evaluation.

Table 8 Speech recognition performance (WER%) per grade class

		1 syllable	2 syllable	3+ syllable
Grade class3	Baseline ASR	38	36	–
	New Algorithm	24	22	–
Grade class 4	Baseline ASR	37	35	11
	New Algorithm	22	20	11
Grade class 5	Baseline ASR	36	34	9
	New Algorithm	20	18	9
Average	Baseline ASR	37	35	10
	New Algorithm	22	20	10

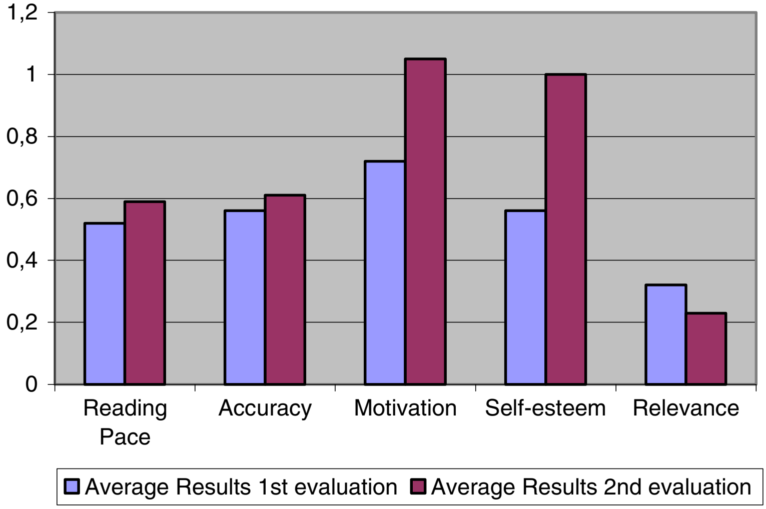

Fig. 5 Children's performance after both evaluations

- Development of a speech recognizer for all words used in the list of sentences (sensitive, alternative, other).

The above list of actions, together with the localization of the graphical user interface (GUI), specifies a tested path in extending the system to new language and the development of the corresponding prototypes.

Besides the system extensibility with respect to the system's target language, the system can also be extended within a single language framework in terms of the number of employed error types (and consequently, the size of the recognized vocabulary).

The Agent-DYSL system is also scalable in terms of the number of dyslexic users that can use it and the complexity and amount of the relevant user data. This scalability is the result of two factors. First, the system is designed to operate in a single-user mode, enabling in that manner its use by any number of users. In addition, the type of data stored for each user in a profile (discrete in nature, non-speech, non-movie like) results in small and manageable prolile files, extending in that sense the time frame that the system can be uses by each user. Second, the expert data knowledge is stored in a scalable knowledge-core component (more powerful than a database system), resulting in a scalable system with respect to the amount of expert data that can be stored into it. As a consequence, the system can practically accommodate any number of error types (and the accompanying expert data).

Another important issue for future work is to part Agent-DYSL onto cost-effective hand-held devices.

Acknowledgments This research work has been supported by the European IST-2005-2.5.11 e-inclusion program within the project "Agent-DYSL", (www.Agent-DYSL.eu).

References

1. AGENT-DYSL, EU project, http://www.Agent-DYSL.eu/
2. AllWrite Company: Sensory Software International Ltd http://www.sensorysoftware.com
3. Amiri H (2006) Reading assistance program for people with dyslexia. PhD thesis, Clayton School of Information Technology, Monash University
4. Athanaselis T, Bakamidis S, Dologlou I, Cowie R, Douglas-Cowie E, Cox C (2005) ASR for emotional speech: clarifying the issues and enhancing performance. Neural Netw Elsevier 18(4):437–444
5. Claro Software Ltd. http://www.clarosoftware.com/
6. Duchateau J, On Kong Y, Cleuren L, Latacz L, Roelens J, Samir A, Demuynck K, Ghesquière P, Verhelst W, Van Hamme H (2009) Developing a reading tutor: design and evaluation of dedicated speech recognition and synthesis modules. Speech Comm
7. Elkind J (1998) Computer reading machines for poor readers, report 9801, Lexia Institute
8. Elkind J, Cohen K, Murray C (1993) Using computer-based readers to improve reading comprehension of students with dyslexia. Ann Dyslexia 42:238–259
9. Fogarty J, Dabbish L, Steck D, Mostow J (2001) Mining a database of reading mistakes: for what should an automated reading tutor listen? 10th Artificial Intelligence in Education
10. Hasselt CV (2002) Oral reading achievements, strategies and personal characteristics of New Zealand primary school students reading below normal expectation. National Education Monitoring Project (NEMP) Probe Study, http://nemp.otago.ac.nz
11. Huslander J (2001) General Information on coding oral reading errors. Olson Reading Lab University of Colorado
12. Kolatch E (2000) Designing for users with cognitive disabilities. College Park, MD: Department of Computer Science White Paper [on-line] http://www.otal.umd.edu/UUGuide/erica/
13. Kurzweil Educational Systems, http://www.kurzweiledu.com/
14. LEXIA READING Sensory Software International Ltd http://www.lexialearning.com/
15. LOGOTYPOGRAFIA. http://www.ilsp.gr/
16. Moore DW, Bean TW, Birdyshaw D, Rycik J (1999) Adolescent literacy: a position statement. International Reading Association
17. Mostow J, Roth S, Hauptmann AG, Kane M (1994) A prototype reading coach that listens. AAAI-94, Seattle, pp 785–792
18. Rack JP (2004) Review of research on effective intervention. In: Turner M, Rack J (eds) The study of dyslexia. Kluwer/Plenum, New York
19. Rapid Readers Company: Rapid http://www.myprimary.co.uk
20. Read Outloud Solo, Don Johnston Special Needs Ltd, http://www.donjohnston.com/
21. ReadOn Company: iTech Consulting Pty Ltd, Australia http://www.readonsoftware.com
22. RocketReader Pty Ltd, http://www.rocketreader.com/
23. Russell M, Brown C, Skilling A, Series R, Wallace J, Bonham B, Barker P (1996) Applications of automatic speech recognition to speech and language development in young children. Proceedings of the International Conference on Spoken Language Processing, ICSLP'96, Philadelphia, USA, 3–6 October 1996
24. Smythe I, Everatt J, Salter R (Eds.) International Book of Dyslexia: a cross-language comparison and practice guide: London John Wiley & Sons
25. Sonant Software, http://www.sonantsoft.com/
26. Special educational needs in Europe "The teaching & learning of languages" http://ec.europa.eu/education/languages/pdf/doc647_en.pdf
27. Tzouveli P, Schmidt A, Schneider M, Symvonis A, Kollias S (2008) Adaptive reading assistance for the inclusion of students with dyslexia: the AGENT-DYSL approach, advanced learning technologies, 2008. ICALT'08, pp. 167–171
28. Williams S, Reiter J (2003) A corpus analysis of discourse relations for natural language generation. Proceedings of Corpus Linguistics 2003 Lancaster University
29. Williams S, Reiter E (2004) Reading errors made by skilled and unskilled readers: evaluating a system that generates reports for people with poor literacy. University of Aberdeen Department of Computing Science Technical Report AUCS/TR0407, pages 1–6
30. Wise B, Cole R, Van Vuuren S, Schwartz S, Snyder L, Ngampatipatpong N, Tuantranont J, Pellom B (2008) Learning to read with a virtual tutor. In: Kinzer C, Verhoeven L (eds) Interactive literacy education: facilitating literacy environments through technology. Lawrence Erlbaum, Taylor and Francis Group, NJ, pp 31–75
31. Young J (1996) Large vocabulary continuous speech recognition. IEEE Signal Process Mag 13(5):45–57
32. Zue V, Seneff S, Polifroni J, Meng H, Glass J (1996) Multilingual human-computer interactions: from information acess to language learning. ICSLP-96, Philadelphia

Dr Theologos Athanaselis received the Diploma degree in Electrical and Computer Engineering from Democritus University of Thrace in September 1999. He holds an Ms'c in Engineering and Physical Sciences in Medicine from the Centre for Biological & Medical Systems of Imperial College of Science, Technology and Medicine. He received his PhD degree in Electrical and Computer Engineering Dept. of NTUA in 2007. The title of PhD dissertation is "Algorithms for application of N-grams in Automatic Spontaneous speech recognition and text processing". His research interests include statistical methods for language processing, spontaneous speech recognition, and speech enhancement techniques. Since 2001, he is working as a Researcher in Voice and Sound Technology Department of Institute for Language and Speech Processing (ILSP/R.C "ATHENA"). Also, he has worked at ILSP/R.C. "Athena" on several projects as a researcher (IST-ERMIS, IST-HEARCOM, IST-AGENT_DYSL, IST-SOPRANO, EHΓ-Aythormito, EHΓ-SUB4ALL).

Stelios Bakamidis received his M.S. degree in Electrical Eng. from the National Technical University of Athens (NTUA) in 1982 (diploma thesis title: "Minimum variance deconvolution of seismic signals"), the MSc degree in Electronics (thesis title: "The Effects of Band-Limiting and Hard-Limiting in Various Modulation Schemes") from the Kings and Chelsea Colleges of the University of London in 1985 and the Ph.D. degree (dissertation title: "Speech Synthesis by Rules for Greek") from the National Technical University of Athens in 1988. From 1986 until 1988 he worked as a teaching assistant at the National Technical University of Athens. From 1989 until 1992 he worked in the private sector. He is researcher A' and Head of the Voice and Sound Technology Department of the Institute for Language and Speech Processing (ILSP).

His research interests are in speech and language processing, pattern recognition and spectral estimation and he has about 60 publications in scientific journals, conferences, books and patents in these fields. Also, he has participated in more than 30 national and EU research projects.

Ioannis Dologlou received his M.S. degree in Electrical Eng. from the National Technical University of Athens (NTUA) in 1981 and the Ph.D. degree from the Institut National Polytechnique of Grenoble in 1984. From 1984 until 1986 he was a postdoctoral fellow at UC Santa Barbara and from 1986 until 1988 he worked as research assistant at Imperial College London. From 1989 until 1991 he was research fellow at NTUA and consultant for the European project in automatic translation EUROTRA-EL. During 1992 he was visiting professor at the University of Paris-Orsay and Paris-Nord and from 1993 until 1994 he was Research Director at the French CNRS. In 1995 and 1996 he was research scientist at the Technical University of Delft and at the Katholieke University of Leuven respectively.

From October 1996 until November 1998 he was the head of the speech group of the EE Dept. of the Katholieke University of Leuven.

Since March 1999 he is a senior researcher at the Institute for Language and Speech Processing (ILSP).

His research interests are in speech and language processing and SVD based spectral estimation and he has over 100 publications in scientific journals and conferences in these fields.

Evmorfia N. Argyriou graduated with distinction from the postgraduate program in 2007. Best postgraduate student of spring semester 2006.

2008-today: Ph.D Candidate in School of Applied Mathematical and Physical Sciences, National Technical University of Athens, Greece

2005-2007: Master in Mathematical Modeling in Modern Technologies and Financial Engineering, National Technical University of Athens, Greece

2000-2005: Diploma in Applied Mathematical and Physical Sciences, National Technical University of Athens, Greece

Antonis Symvonis primary research interests are in the theoretical aspects of Computer Science, focusing on algorithms and their complexity (graph drawing, packet routing, parallel computation, graph theory, computational geometry).

Ph.D in Computer Science, University of Texas at Dallas, USA, 1991

Master in Computer Science, University of Texas at Dallas, USA, 1989

Diploma in Computer Engineering and Information Sciences, University of Patras, Greece, 1987