a) Integrated I/O Environment

5.1.1 VOICE - A Voice Oriented Interactive Computing Environment

R. Ahuja, Bhiksha Raj, N. Bondale, X. Furtado, Thomas Jose, S. Krishnan, P. Poddar, P.V.S. Rao, K. Samudravijaya and A. Sen, *Special issue on the KBCS Project, Vivek, Vol. 5, No. 4, Oct. 1992.*

Abstract

The objective of the project was to develop an input/ output system to a computer with a facility for visual and voice feedback. Using primarily the speech mode, this was to be an interactive facility with provisions for keyboard entry, voice output and visual display. The aim was to make it possible for uninitiated users to interact with computers. Considering that there was no earlier work on speech in Indian languages a system working in a well defined and strictly delimited task environment was aimed at. It was to be speaker dependent with a vocabulary of about 200 words. The system was planned to accept clearly spoken isolated/ connected words and to produce intelligible speech.

With the aim of realizing the above goals, several outputs were delivered at various stages of the project. The paper contains a description of major accomplishments with brief technical background.

5.1.2 OCR And Speech Recognition For Oriya Language

Sanghamitra Mohanty, *www.emille.lancs.ac.uk/ lesal/mohanty.pdf.*

Abstract

Optical Recognition System and Speech Recognition system are the technology part, which can be of immediate help to a person without deep knowledge of the language or English. We have used the Neural Network technique in both the cases for making it noise free and also for the recognition. Our effort is towards the integration of the OCR system to speech Recognition System to make effective use of both.

5.1.3 Adapting Question Answering Techniques To The Web

Jignashu Parikh, M. Narasimha Murty, *Language Engineering Conference, University of Hyderabad, India, Dec. 2002.*

Abstract

The Web has emerged as a huge information repository that can be used for various knowledge based applications, an important one being Question Answering (QA). The paper discusses the issues involved in using web as a knowledge base for question answering involving simple factual questions. It also proposes some simple but effective methods to adapt traditional QA methods to deal with these issues and lead to an accurate and efficient Question Answering system. The paper discusses the architecture and design of our system. The empirical results obtained on the current system are encouraging.

5.1.4 Foresight - A Personalised Syntax Based Prediction System for Natural Language Text Input into Computers

P.V.S. Rao, *Gabungan Komputer Nasional Malaysia and Management Science/Operations Research Society of Malaysia (MSORSM), ICIT'90, Sept. 17-19, 1990, Kuala Lumpur.*

Abstract

Most users of computers who are required to do text entry in natural languages have little or no experience in touch-typing. They are therefore slow and quite reluctant to use the machine. The present paper describes an effort to improve this situation by very significantly reducing the number of key strokes required to text entry.

The first level of prediction is based on a dictionary search. For any reasonable vocabulary size, the word to be typed in can be determined as soon as an adequate number of characters (say three or four) have been typed in. The word "algebra" can, for instance be guessed as soon as the sequence /alg/ has been typed. The syntax of the language can help to improve this performance further. The words in the sentence "I am very grateful to you for your letter" can, for instance, be guessed with increasing certainty as more and more of the sentence is typed. This prediction can be made using one of two models of the syntax: a word n-gram model and a stochastic probabilistic network model. One can build up either a general model for the language or a personalized model for the writer himself.

Key strokes can be reduced by 50% to 70% using even very simple systems. Further work to refine the system is continuing and is expected not only improve the efficiency of text entry but to help understand the strengths and weaknesses of alternate language models. There are other applications for such models in the areas of speech and connected script recognition, settling authorship disputes by stylistic analysis, etc.

5.1.5 NLP Research In The CSR Lab Of Tata Infotech

P.V.S Rao, *2nd Annual Convention of the North American Chapter of the Association for Computational Linguistics, Pittsburgh, June 6, 2001*

Abstract

OASIS is an interactive question - answering system, which answers questions in common English. Developed as a result of on-going research at the Tata Infotech's, Cognitive Systems Research Lab, OASIS is specifically oriented towards answering questions and providing information on a given topic, and is the first product of this type to be developed in India. The engine has been set-up to answer questions about Tata Infotech.

The system adopts a user-friendly conversational style. Each answer is preceded by a conversational 'prelude' which confirms to the user that the question has been properly understood. It also suggests other topics on which the user can ask questions. In its present form, OASIS will not deal with involved questions of certain types and extremely complicated sentences.

A variant of this has been implemented which answers questions relating to textual information as well as information from standard databases. The question in natural language is in this case converted into a query in SQL and the information obtained from the database is presented to the user, embedded in an English sentence. OASIS is proposed to be made available both as a product and as a solution

b) Optical Character Recognition (OCR)

5.1.6 A Hybrid Scheme For Handprinted Numeral Recognition Based On A Self-Organizing Network And MLP-Based Classifiers

U. Bhattacharya, T. K. Das, A. Datta, S. K. Parui and B. B. Chaudhuri, *International Journal of Pattern Recognition & Artificial Intelegence, vol 16, no 7 (2002) pp 845-864.*

Abstract

This paper proposes a novel approach for automatic recognition of handprinted Bangla (an Indian script) numerals. A Topology Adaptive Self Organizing Neural Network is first used to extract vector skeleton from a binary numeral pattern. Simple heuristics are considered to prune artifacts, if appeared, in such a skeletal shape. Certain structural and topological features like loops, junctions, positions of terminal nodes etc. are used along with a hierarchical tree classifier to classify handwritten numerals into smaller subgroups. A multilayer perceptron network (MLP) is used to classify different numerals in each subgroup uniquely. The system is trained using a sample data set of 1800 numerals and we obtained 93.26% correct recognition rate and 1.71% rejection on a disjoint test set of another 7760 samples. In addition, a separate validation set consisting of 1540 samples has been used to determine the termination of the training of the associated MLP networks. The proposed scheme is sufficiently robust with respect to considerable object noise and it does not consider any size normalization.

5.1.7 On Developing High Accuracy OCR Systems For Telugu And Other Indian Scripts

Chakravarthy Bhagavati, Tanuku Ravi, S. Mahesh Kumar, Atul Negi, *Language Engineering Conference, University of Hyderabad, India, Dec. 2002.*

Abstract

In this paper, we list a number of factors that are important in achieving high recognition accuracy in OCR systems for Telugu and other Indian scripts. While it is relatively easy to obtain 85% - 93% accuracy, it becomes increasingly difficult to improve the performance further. We discuss how the factors presented in this paper helped achieve an accuracy of nearly 97% with our OCR system for Telugu script. It is expected that these factors are specific not only to Telugu but also work for other Indian scripts in general and south Indian scripts in particular.

5.1.8 Recognition Of Handprinted Bangla Numerals Using Neural Network Models

U. Bhattacharya, T. K. Das, A. Datta, S. K. Parui, and B. B. Chaudhuri, *Advances in Soft Computing - AFSS 2002, Springer Verlag, Lecture Notes on Artificial Intelligence, Eds. N.R. Pal and M. Sugeno, LNAI 2275, 2002, pp. 228-235.*

Abstract

This paper proposes an automatic recognition scheme for handprinted Bangla (an Indian script) numerals using neural network models. A Topology Adaptive Self Organizing Neural Network is first used to extract from a numeral pattern a skeletal shape that is represented as a graph. Certain features like loops, junctions etc. present in the graph are considered to classify a numeral into a smaller group. If the group is a singleton, the recognition is done. Otherwise, multilayer perceptron networks are used to classify different numerals uniquely. The system is trained using a sample data set of 1880 numerals and we obtained 90.56% correct recognition rate on a test set of another 3440 samples. The proposed scheme is sufficiently robust with respect to considerable object noise.

5.1.9 Self-Organizing Neural Network-Based System For Recognition Of Handprinted Bangla Numerals

U. Bhattacharya, T. K. Das, A. Datta, S. K. Parui, and B. B. Chaudhuri, *Proceedings of XXXVI Annual Convention, Computer society of India, 2001, Kolkata, pp. C-92 - C-96.*

Abstract

This paper proposes an automatic recognition scheme for handprinted Bangla (an Indian script) numerals using a Topology Adaptive Self Organizing Neural Network (TASONN) model. The Neural Network model is used to extract from a numeral pattern a skeletal shape that is represented as a planar straight line graph. Certain features like loops, junctions etc. present in the graph are considered to classify a numeral into smaller groups. If the group is a singleton, the recognition is done. Otherwise the graph is subjected to various feature extraction procedure depending on the group. Recognition is done using a look-up table of specific ranges of the extracted feature values. This scheme does not require normalization with respect to size and it performs satisfactorily even in the presence of considerable noise. The system tested on a test set of 3330 samples with 89.63% correct recognition rate.

5.1.10 A Complete Printed Bangla OCR System

B. B. Chaudhuri and U. Pal, *Pattern Recognition, vol. 31, pp. 531-549, 1998.*

Abstract

A complete Optical Character Recognition (OCR) system for printed Bangla, the fourth most popular script in the world, is presented. This is the first OCR system among all script forms used in Indian sub-continent. The problem is difficult because (i) there are about 300 basic, modified and compound character shapes in the script, (ii) the characters in a word are topologically connected and (iii) Bangla is an inflectional language. In our system the document image captured by Flat-bed scanner is subject to skew correction, text graphics separation, line segmentation, zone detection, word and character segmentation using some conventional and some newly developed techniques. From zonal information and shape characteristics, the basic, modified and compound characters are separated for the convenience of classification. The basic and modified characters which are about 75 in number and which occupy about 96% of the text corpus, are recognized by a structural feature based tree classifier. The compound characters are recognized by a tree classifier followed by template matching approach. The feature detection is simple and robust where preprocessing like thinning and pruning are avoided. The character unigram statistics is used to make the tree classifier efficient. Several heuristics are also used to speed up the template matching approach. A dictionary based error correction scheme has been used where separate dictionaries are compiled for root word and suffixes that contain morpho-syntactic information's as well. For single font clear documents 95.5% word level (which is equivalent to 99.10% character level) recognition accuracy has been obtained. Extension of the work to Devnagari, the third most popular script in the world, is also discussed.

5.1.11 An OCR System To Read Two Indian Language Scripts: Bangla And Devnagari (Hindi)

B. B. Chaudhuri and U. Pal, *Proc. Fourth Int. conf. on Document Analysis and Recognition, IEEE Computer Society Press, pp. 1011-1016, 1997.*

Abstract

An OCR system is proposed that can read two Indian language scripts: Bangla and Devnagari (Hindi), the most popular ones in Indian subcontinent. These scripts, having the same origin in ancient Brahmi script, have many features in common and hence a single system can be modeled to recognize them. In the proposed model, document digitization, skew detection, text line segmentation and zone separation, word and character segmentation, character category are done for both scripts by the same set of algorithms. The feature sets and classification tree as well as knowledge base required for error correction (such as lexicon) differ for Bangla and Devnagari. The system shows a good performance for single font scripts printed on clear document.

5.1.12 Automatic Recognition Of Printed Oriya Script

B. B. Chaudhuri, U. Pal and M. Mitra, Sadhana, *(a journal of Indian Academy of Sciences) vol.27, part 1. pp.23-34, 2002.*

Abstract

This paper deals with an Optical Character Recognition (OCR) system for printed Oriya script. The development of OCR for this script is difficult because a large number of character shapes in the script have to be recognized. In the proposed system, the document image is first captured using a flat-bed scanner and then passed through different preprocessing modules like skew correction, line segmentation, zone detection, word and character segmentation, etc. These modules have been developed by combining some conventional techniques with some newly proposed ones. Next, individual characters are recognized using a combination of stroke and run-number based features, along with features obtained from the concept of water overflow from a reservoir. The feature detection methods are simple and robust, and do not require preprocessing steps like thinning and pruning. A prototype of the system has been tested on a variety of printed Oriya material, and currently achieves 96.3% character level accuracy on average.

5.1.13 Development Of A Page Layout Analyzer For Multilingual Indian Documents

Ray Chaudhuri, A.K.Mandal, B.B.Chaudhuri *(Fellow IEEE), Language Engineering Conference, University of Hyderabad, India, Dec. 2002.*

Abstract

An advanced Optical Character Recognition (OCR) system is equipped the module of the page layout analyser. It separates textual zones from non-textual zones. It identifies textual blocks from multicolumn documents and groups them into homogenous regions in terms of geometric shape and spatial distribution. All existing OCR modules developed for various Indian scripts can handle text only single-column documents. In this paper, a page layout analyser that uses typical common features present in most of the Indian scripts is introduced. A simple compatibility criterion that allows various degrees of homogeneity is defined. The page-analyser is robust in the sense that it can distinguish text regions from non-textual entities such as images, rulers, and noisy signals due to smudges and poor quality of the paper. Test results are shown in two most popular Indian Scripts, Devnagari (Hindi) and Bangla.

5.1.14 Skew Angle Detection Of Digitized Indian Script Documents

B.B. Chaudhuri and U. Pal, *IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, pp.182-186, 1997.*

Abstract

Skew angle detection of scanned documents containing most popular Indian scripts Devnagari and Bangla is considered. Most characters in these scripts have horizontal lines at the top, called head lines. The character head lines mostly join one another in a word and the word appears as a single component. In the proposed method the components are at first labeled. The upper envelope of a component is found by column-wise scanning from an imaginary line above the component. Portions of upper envelope satisfying the properties of digital straight line are detected. They are clustered as belonging to single text lines. Estimates from individual clusters are combined to get the skew angle. Apart from accuracy and efficiency, an advantage of the method is that character segmentation and zone detection can be readily done from head line information, which is useful in Optical Character Recognition approaches of these scripts.

5.1.15 Optical Character Recognition (OCR) System For Malayalam Language

K. Jithesh, K.G. Sulochana, R Ravindra Kumar *National Workshop on Application of Language Technology in Indian Languages, to be held in Hyderabad, March 6-8, 2003*

Abstract

In this paper we present a brief description of a Malayalam Optical Character Recognition (OCR) system. The presence of two different scripts (Old Script and New Script) and a large number of characters (including conjuncts) makes the Malayalam OCR a complex system. The proposed system is based on the **Feature Extraction** method of character recognition. Feature extraction can be considered as finding a set of vectors, which effectively represent the information content of a character. In character recognition, it is desirable to extract features, which are focused on discriminating between classes. The features are identified after the careful study of Malayalam writing system. A two level segmentation scheme, feature extraction method and classification scheme, using binary decision tree, for Malayalam characters is described. Different pre-processing modules like noise removal, skew detection and correction, line, word and character segmentation are also dealt with. The presence of touching characters particularly in the case of consonant-vowel modifier combinations renders the character segmentation process difficult. The development of the post processor, a spell checker tuned for the OCR system, is in progress. The objective of the post processing is to correct errors in OCR output by using Malayalam grammar rules, lookup table (word list) and statistical information collected from the corpora. The beta version of the System gives an accuracy of 97% at character level for good quality printouts. We are now working on Document layout analyser module, the addition of which will enable the OCR to reproduce the document in its original layout.

5.1.16 A Complete OCR System For Gurmukhi Script

G S Lehal and Chandan Singh, *Proceedings SSPR2002, Windsor, Canada, Lecture Notes in Computer Science, Vol. 2248, Springer-Verlag, Germany, pp. 344-352, (2002)*

Abstract

Recognition of Indian language scripts is a challenging problem. Work for the development of complete OCR systems for Indian language scripts is still in infancy. Research in the field of recognition of Gurmukhi script faces major problems mainly related to the unique characteristics of the script like connectivity of characters on the headline, characters in a word present in both horizontal and vertical directions, two or more characters in a word having intersecting minimum bounding rectangles along horizontal direction, existence of a large set of visually similar character pairs, multi-component characters, touching characters which are present even in clean documents and horizontally overlapping text segments. This paper

addresses the problems in the various stages of the development of a complete OCR for Gurmukhi script and discusses potential solutions. A multi-font Gurmukhi OCR for printed text with an accuracy of more than 97% at character level is presented. The recognition system presented in this paper operates at connected component level. The segmentation process decomposes the text image into connected components. After feature extraction, the connected components are fed to a classifier, which recognizes the connected component. The connected components are then combined to form Gurmukhi characters. A set of very simple and easy to compute structural features is used and a hybrid classification scheme consisting of binary decision trees and nearest neighbours is employed. To further improve the results at word level, each word is fed to a post processor, which uses the statistical information of Punjabi language syllable combination, corpora look up and certain heuristics based on Punjabi grammar rules.

5.1.17 A Range Free Skew Detection Technique For Digitized Gurmukhi Script Documents

G S Lehal and Renu Dhir, *Proceedings 5th International Conference of Document Analysis and Recognition, Bangalore, pp. 147-152, (1999)*

Abstract

In this paper, a range free skew detection technique for machine printed Gurmukhi documents has been presented. Most characters in Gurmukhi script have horizontal lines at the top called headlines. The characters forming a word are joined at top by headlines, so that the word appears as one single component with headline. The ratio of pixel density above and below the headline of any word in Gurmukhi script is always less than one. These inherent characteristics of the script have been employed and a new algorithm based on projection profile method has been devised. The skew angle is determined by calculating horizontal and vertical projections at different angles at fixed interval in range [0°, 90°]. Under such projections, for an image with no skew, headlines appear as distinct peaks while gaps between successive text rows will be represented by valleys. The bitmapped image is partitioned into ten equal sized horizontal and vertical zones and the highest peaks and valleys are determined for projections in each zone. The angle at which the difference of the sum of heights of peaks and valleys is maximum is identified as the skew angle. To decrease the computational cost, first the course skew angle is calculated by taking the angle interval 3°. Once the course skew angle is found, the accurate skew angle q is determined by looking in the range [q - 3°, q + 3°] at intervals of 0.25°. The image is then rotated over -q, where q is the skew angle

Since the skew angle is checked only in the range [0°-90°] and the image can be skewed at any angle in the range [-180°, 180°], the rotated image may need another additional rotation by 90°, -90° or 180°. If the rotated image is skewed at 90° or -90°, then the highest peaks and valleys would be present in vertical projection else they will be reported in horizontal projection. To determine the skew angle of the image aligned with y-axis, if the foreground pixel density on the left side of headlines is greater than pixel density on right side for text rows then the image is skewed at -90 else it is skewed at 90°. Similarly for the image aligned with x-axis, if the foreground pixel density above the headlines is lesser than pixel density below then the image is straight else it is upside down. In the end the image is rotated by the second rotation angle to completely remove any skew present in the image.

5.1.18 A Recognition System For Devnagri And English Handwritten Numerals

G. S. Lehal and Nivedan Bhatt, *Advances in Multimodal Interfaces – ICMI 2001, T. Tan, Y. Shi and W. Gao (Editors), Lecture Notes in Computer Science, Vol. 1948, Springer-Verlag, Germany, pp. 442-449. (2000).*

Abstract

Handwritten numeral recognition has been extensively studied for many years and a number of techniques have been proposed [1-5]. However, handwritten character recognition is still a difficult task in which human beings perform much better. The problem of automatic recognition of handwritten bilingual numerals is even more tough. In this paper a bilingual OCR system for handwritten numerals of Devnagri(Hindi) and Roman scripts has been presented. It is assumed at a time the numerals will be of one of the above two scripts and there are no mixed script numerals in an input string. A set of global and local features, which are derived from the right and left projection profiles of the numeral image, are used. For identification of script, no separate routine is used. During the recognition process when the first numeral of a numeric string is recognized correctly, the context (Devnagri/ English) is set to the domain of that particular numeral's character set. Subsequent identification of the remaining numerals is carried out in that context only which drastically reduces the search space and hence increases the performance of the system. It was observed that the Devnagri numeral set had a very good recognition and rejection rate, as compared to the English set. Also the Devnagri numeral set's recognition module had good rejection rates for the numerals of the English character set. This property was exploited by adopting a *polling strategy* in which the input numeral is first tested by the Devnagri module.

If the numeral is recognized then the context is set to Devnagri, else it is tested for English set and on recognition, the context is set to English. In case the numeral is rejected by both script sets, then the next numeral is tested, and this continues till one of the numeral is recognized. Subsequent identification of the other numerals is carried out for character set of the recognized numeral. Numeral 0 is the same for both the character sets, thus in the case when the first numeral encountered is a zero, subsequent numeral is checked before deciding the context. The system was tested on 1000 samples of both the Devnagri and English character set. For the Devnagri numeral set, a recognition rate of 89% and a confusion rate of 4.5% were obtained. For the English numeral set we had a recognition rate of 78.4%, confusion rate of 18 % and rejection rate of 3.6%.

5.1.19 A Shape Based Post Processor For Gurmukhi OCR

G S Lehal, Chandan Singh and Ritu Lehal, *Proceedings of 6th International Conference on Document Analysis and Recognition, Seattle, USA, IEEE Computer Society Press, USA, pp. 1105-1109, (2001)*

Abstract

The objective of post processing is to correct errors or resolve ambiguities in OCR results by using contextual information. In this paper a shape based post processing system for an OCR of Gurmukhi script has been presented. The Punjabi corpus developed at MIT under TDIL project has been used, which serves the dual purpose of providing data for statistical analysis of Punjabi language and also checking the spelling of a word. The corpus has been partitioned at two levels. At the first level the corpus is split into seven disjoint subsets based on the word length. At second level the shape of the word is used to further segment the subset into a list of visually similar words. A set of robust, font and character size independent features are used for identification of visually similar words. These features are available more or less as a by-product of the on-going recognition process and do not necessitate any additional computation. For each set of visually similar word the percentage frequency of occurrence of character in all the positions is recorded. This list is combined with the confidence rate of recognition of the recognizer to correct the mistakes of the recognizer. Holistic recognition of most commonly occurring words derived from the corpora is also carried out. An improvement of 3% in recognition rate from 94.35% to 97.34% has been reported on machine printed images using the post processing techniques.

5.1.20 A Structural Feature Based Approach For Script Identification Of Gurmukhi And Roman Characters And Words

G S Lehal, Chandan Singh and Renu Dhir, *Proceedings SPIE, (Jan. 2003)*

Abstract

Roman script words are now commonly being used in Gurmukhi script documents. An OCR developed for the Gurmukhi script will wrongly recognize these words in Roman script. So it is necessary to filter out these Roman script words before feeding the Gurmukhi script words to the OCR. Considering the nature of many documents in the Indian context, where the script could change at the word level or even single characters of different script may be present, there is need for development of method to identify the script of a character or a word. In this paper we have proposed a method to automatically differentiate between Gurmukhi and Roman script words and characters based on a combined analysis of several discriminating features. After a careful study of shapes of Gurmukhi and Roman script characters and words we have developed nine features for automatic classification of Roman and Gurmukhi scripts. Some of these features are common for identification at both character and word level, while some features are suitable only for either word or single characters only. These features are *Headline pixel count , Inter character gap, Bottom Projection Profile, Protruding Regions Beyond Headline, Right Vertical Bar, Loop in lower half, Left Vertical bar, C shape in lower half* and *U like shape*. This method has been implemented and tested on about 100 documents, and the experimental results indicate that this method is effective and reliable. The recognition accuracy of the system is 98.29% for Gurmukhi script words and 99.02% for Roman script words. For single characters, the accuracy is 96.81% and 95.47% for Roman and Gurmukhi scripts respectively. This is the first time that such a script recognition system has been developed for Roman and an Indian language script, which works down to word and character level.

5.1.21 A Technique For Segmentation Of Gurmukhi Text

***G S Lehal and Chandan Singh**, Computer Analysis of Images and Patterns,* W. Skarbek *(Ed.), Lecture Notes in Computer Science, Vol. 2124, Springer-Verlag, Germany, pp. 191-200 (2001)*

Abstract

In this paper a technique for text segmentation of machine printed Gurmukhi script documents is discussed. Research in the field of segmentation of Gurmukhi script faces major problems mainly related to the unique characteristics of the script like

connectivity of characters on the headline, two or more characters in a word having intersecting minimum bounding rectangles, multi-component characters, touching characters which are present even in clean documents and horizontally overlapping text segments. After digitization of the text, the text image is subjected to pre-processing routines such as noise removal, thinning and skew correction. The thinned and cleaned text image is then sent to the text segmenter, which segments the text image into connected components. The text image is sliced into horizontal text strips using horizontal projection in each row. The gaps on the horizontal projection profile are taken as separators between the text strips. But this step does not always results in a single text row in each horizontal strip. Usually a text line is broken up into two or more horizontal strips. We call these strips as zones. The text line is broken into a single core zone, which is made up of characters of middle and upper zone and optionally lower zones, followed by other minor zones representing the lower zones. In some rare cases the core zone is also split into two zones representing the upper and middle zones. Some other problems also occur such as the lower characters and vowels of some of the text lines intruding into the core zone of the successor text line and more than one text line being present in a horizontal strip because of overlap. A statistical analysis of strip heights and position of headline is used to identify if a strip contains more than one text line or only a lower zone or middle and upper zone or only upper zone of a text line. The text strips are next decomposed into connected components. In case of multi strips, which contain multiple text lines, horizontal cuts are made at appropriate locations and the connected components are located in each horizontal cut. Simple heuristics are also used to detect and split touching connected components in upper and middle zones.

5.1.22 Feature Extraction And Classification For OCR Of Gurmukhi Script

G S Lehal and Chandan Singh, Vivek, *Vol. 12, No. 2, pp. 2-12 (1999).*

Abstract

In recent years, there has been a renewed attempt to reformat the classification approaches to the recognition of difficult character sets. It has been found that a multiple classification character recognition scheme has the potential of outperforming individual stand-alone classifiers because of its ability to handle extreme variance in the training and testing samples. In this paper, a feature extraction and a hybrid classification scheme, using binary decision trees and nearest neighbours, for machine recognition of Gurmukhi characters is described. The classification process is carried out in three stages. In the first stage the characters are grouped into three sets depending on their zonal position (upper zone set, middle zone set and lower zone). In the second stage the characters in middle zone set are further distributed into smaller sub-sets by a binary decision tree using a set of robust and font independent features at each node of the tree. The terminal node of the binary decision tree contains a subset of Gurmukhi characters and the cardinality of the set varies from 1 to 8. The final categorization of the input sample is then easily tackled by using a suitable recognition schemes for each subset, considering the special features and peculiarities of the characters in each subset. In the third stage the nearest neighbour classifier is used and the special features distinguishing the characters in each subset are used. One significant point in this scheme is that in contrast to the conventional single-stage classifiers where each character image is tested against all prototypes, a character image is tested against only certain subsets of classes at each stage. This eliminates unnecessary computations.

5.1.23 A New Algorithm For The Restoration Of Characters In Old Noisy Document With Varying Level Of Intensities

S. Mohanty, K. Sahoo and H. K. Behera, *Indian Science Congress, 2003.*

Abstract

An image can have noise and interference from several sources like electrical sensor, photographic grain, and channel errors. Noise cleaning is the process of removing unwanted noise from an image and is necessary for better recognition of characters in a document. These noise effects must be reduced for better analysis of character documents, which will lead to efficient character extraction and recognition process. We have developed a novel algorithm, which is simple and clears noise from character document images to a very good extent. In the proposed methodology, the algorithm is based on intensities, which clears the noises and isolated points present on the document.

5.1.24 A Solution For Ligatures During Optical Recognition of Oriya Characters

S. Mohanty, K. Sahoo and H. K. Behera, *Published at: Proceedings of IEMCT 2002, CDAC, Pune.*

Abstract

Recognition of alphabetic characters is a basic need in incorporating intelligence to computers. Machine intelligence involves several aspects among which optical recognition is a tool, which can be integrated to text recognition and speech recognition. To make these aspects effective character recognition with better accuracy is important.

Very often even in printed text, adjacent characters tend to be touched or connected. Generally, through binarization of images, useful information for the segmentation of touched character are lost in many cases which leads to the accuracy being handicapped when problem like ligatures predominates during recognition. We in our study have developed an algorithm, which provides a solution to overcome the complexity of ligatures especially for connected characters during segmentation leading towards more accurate recognition.

From the Histogram of different size neighborhoods of characters the optimal character size found out and tested on further segmentation. This technique is a recognition-based segmentation for connected characters.

The algorithm has successfully been tested for Oriya characters and the output has been integrated with Oriya Text-to-Speech system.. This technique can be applied for other Indian as well as some foreign languages during Optical Character Recognition.

5.1.25 Pattern Recognition In Alphabets Of Oriya Language Using Kohonen Neural Network

S. Mohanty, *Published at International Journal of Pattern Recognition and Artificial Intelligence Vol.12 No.7 (1998).*

Abstract

Here a computerized reading of alphabets of Oriya language is attempted using the Kohonen neural network and its unsupervised competitive learning capacity as self- organizing map or the Kohonen feature map. The proposed pattern recognition does not treat a pattern as an n-dimensional feature vector or a point in n-dimensional space as is done in the traditional pattern recognition theory. We have tried with all the Oriya alphabets and have presented the study with respect to five of them in this paper along with their average distance per pattern in each cycle till we reach the permissible average distance. In the output picture the variation of the weight vector with respect to the alphabets is clearly observed.

5.1.26 Page Layout Analyzer For Multilingual Indian Documents

A.Mandal and Prof. B. B. Chaudhuri, *that will be published in the Proceedings of the Language Engineering Conference 2002 by IEEE CS Press.*

Abstract

An advanced Optical Character Recognition (OCR) system is equipped the module of the page layout analyser. It separates textual zones from non-textual zones. It identifies textual blocks from multicolumn documents and groups them into homogenous regions in terms of geometric shape and spatial distribution. All existing OCR modules developed for various Indian scripts can handle text only single-column documents. In this paper, a page layout analyser that uses typical common features present in most of the Indian scripts is introduced. A simple compatibility criterion that allows various degrees of homogeneity is defined. The page-analyser is robust in the sense that it can distinguish text regions from non-textual entities such as images, rulers, and noisy signals due to smudges and poor quality of the paper. Test results are shown in two most popular Indian Scripts, Devnagari (Hindi) and Bangla.

5.1.27 Automatic Identification of English, Chinese, Arabic, Devnagari And Bangla Script Line

U. Pal and B. B. Chaudhuri, *In Proc. Sixth Int. Conf. on Document Analysis and Recognition, IEEE Computer Society Press, pp.790-794, 2001.*

Abstract

In a general situation, a document page may contain several script forms. For Optical Character Recognition (OCR) of such a document page, it is necessary to separate the scripts before feeding them to their individual OCR systems. In this paper, an automatic technique for the identification of printed Roman, Chinese, Arabic, Devnagari and Bangla text lines from a single document is proposed. Shape based features, statistical features and some features obtained from the concept of water reservoir, have been used for script identification. The proposed scheme has an accuracy of about 97.33%.

5.1.28 Automatic Recognition Of Unconstrained Off-Line Bangla Hand-Written Numerals

U. Pal and B. B. Chaudhuri, *Advances in Multimodal Interfaces, Springer Verlag Lecture Notes on Computer Science (LNCS-1948), Eds. T. Tan, Y. Shi and W. Gao, pp. 371-378, 2000.*

Abstract

This paper deals with an automatic recognition method for unconstrained off-line Bangla handwritten numerals. To take care of variability involved in the writing style of different individuals a robust scheme is presented here. The scheme is based on new features obtained from the concept of water overflow from the reservoir as well as topological and statistical features of the numerals. If we pour water from upper part of the character, the region where water will be stored in the character is imagined as a reservoir of the character. The direction of water overflow, height of water level when water overflows from the reservoir, position of the reservoir with respect to the character bounding box, shape of the reservoir etc. are used in the recognition scheme. The proposed scheme is tested on data collected from different individuals of various background and we obtained an overall recognition accuracy of about 91.98%.

5.1.29 Automatic Separation Of Words In Indian Multi-Lingual Multi-Script Documents

U. Pal and B. B. Chaudhuri, *Proc. Fourth Int. Conf. on Document Analysis and Recognition, IEEE Computer Society Press, pp. 576-579, 1997.*

Abstract

In a multi-lingual country like India, a document page may contain more than one script forms. For such a document it is necessary to separate different script forms before feeding them to OCRs of individual script. In this paper an automatic word segmentation approach is described which can separate Roman, Bangla and Devnagari scripts present in a single document. The approach has a tree structure where at first Roman script words are separated using the 'headline' feature. The headline is common in Bangla and Devnagari but absent in Roman. Next, Bangla and Devnagari words are separated using some finer characteristics of the character set although recognition of individual character is avoided. At present, the system has an overall accuracy of 96.09%.

5.1.30 Machine-Printed And Hand-Written Text Lines Indentification

U. Pal and B. B. Chaudhuri, *Pattern Recognition Letters, Vol.22, No. 3-4, pp. 431-441, 2001*

Abstract

There are many types of documents where machine-printed and hand-written texts intermixedly appear. Since the optical character recognition (OCR) methodologies for machine-printed and hand-written texts are different, to achieve optimal performance it is necessary to separate these two types of text before feeding them to their respective OCR systems. In this paper, we present a machine-printed and hand-written text classification scheme for Bangla and Devnagari, the two most popular Indian scripts. The scheme is based on the structural and statistical features of the machine-printed and hand-written text lines. The classification scheme has an accuracy of 98.6%.

5.1.31 Multi-Skew Detection Of Indian Script Documents

U. Pal, M. Mitra and B. B. Chaudhuri, *In Proc. Sixth Int. Conf. on Document Analysis and Recognition, IEEE Computer Society Press, pp. 292-296, 2001.*

Abstract

There are many documents where text lines are not parallel to each other i.e. these lines have different inclinations with the horizontal lines (multi-skew documents). For the OCR of such a document we have to estimate the skew angle of individual text lines because a single rotation cannot de-skew all text lines of the document. In this paper, we describe a robust technique for multi-skew angle detection from Indian documents containing the most popular Indian scripts Devnagari and Bangla. Most characters in these scripts have horizontal lines at the top, called head-lines. The character head-lines usually connect one another in a word and the word appears as a single component. In the proposed method, the connected components are at first labeled and selected. The upper envelopes of selected components are found by column-wise scanning from the top of the component. Portions of the upper envelope satisfying the properties of a digital straight line are detected. They are then clustered into groups belonging to single text lines. Estimates from these individual clusters give the skew angle of each text line. The proposed multi-skew detection technique has an accuracy about 98.3%.

5.1.32 OCR Error Correction Of An Inflectional Indian Language Using Morphological Parsing

U. Pal, P. K. Kundu and B. B. Chaudhuri, *Journal of Information Science and Engineering, Vol.. 16. No.6. pp. 903-922, 2000.*

Abstract

This paper deals with an OCR (Optical Character Recognition) error detection and correction technique for a highly inflectional Indian language, Bangla, the second-most popular language in India and fifth-most popular language in the world. The technique is based on morphological parsing where using two separate lexicons of root words and suffixes, the candidate root-suffix pairs of each input string are detected, their grammatical agreement are tested and the root/suffix part in which the error has occurred is noted. The correction is made on the corresponding error part of the input string by a fast dictionary access technique. To do so, the information about the error patterns generated by the OCR system are examined and some alternative strings are generated for an erroneous word. Among the alternative strings, those satisfying grammatical agreement in root and suffix are finally chosen as suggested words. In the list of suggested words generated by the system, the desired word is available in 84.22% cases.

5.1.33 OCR In Bangla : An Indo-Bangladeshi Language

U. Pal and B. B. Chaudhuri, *Proc. of 12th Int. Conf. on Pattern Recognition, IEEE Computer Society Press, pp. 269-274, 1994.*

Abstract

In this paper a complete OCR system is described for documents of single Bangla (Bengali) font. The character shapes are recognized by a combination of template and feature matching approach. Images digitized by flatbed scanner are subjected to skew correction, line word and character segmentation, simple and compound character separation feature extraction and finally character recognition. A Feature based tree classifier is used for simple character recognition. Preprocessing like thinning and skeletonization is not necessary in our scheme and hence the system is quite fast. At present, system has an accuracy of about 96%. Also, some character occurrence statistics have been computed to model an error detection and correction technique in near future.

5.1.34 On The Development of An Optical Character Recognition (OCR) System For Printed Bangla Script

U. Pal, *Ph.D. Thesis, Indian Statistical Institute, 1997.*

Abstract

This thesis is devoted to the development of a complete OCR system for printed *Bangla* script. The content of the thesis is divided into three major divisions. They are (a) Preprocessing Division (b) Recognition Division (c) Postprocessing Division.

In Preprocessing Division, at first, some statistical studies have been made on Bangla script characters. Also, the application potentials of these studies in OCR design have been described. Next, binarization, smoothing, skew detection and correction as well as text/graphics separation techniques have been elaborated. The skew correction method proposed here is simple, fast and robust. The method has been developed using inherent characteristics of the script form. Finally, techniques for segmentation of text into lines, zone detection, word segmentation from line and character segmentation from word have been described.

In recognition division at first feature selection and detection procedure as well as analysis on binary tree classifier have been presented. Next, automatic character recognition procedure using a hybrid method is described. Combination of a feature based tree classifier and a run length based template matching approach has been used for the purpose. For convenience of classification we have initially classified the characters into three categories namely basic, modified and compound character. The basic and modified class of characters have been recognized by a feature based tree classifier while the compound character recognition involves a template matching approach preceded by a feature based subgrouping.

In postprocessing division we proposed an OCR error detection and correction technique for a highly inflectional language like Bangla. Using two separate lexicons of root words and suffixes, we detect candidate root-suffix pair of each word and test their grammatical agreement, and note the root suffix part in which the error has occurred. The correction is done on the corresponding error part of the input string by a fast dictionary access technique.

5.1.35 Printed Devnagari Script OCR System

U. Pal and B. B. Chaudhuri, Vivek, *vol.10, pp.12-24, 1997.*

Abstract

An Optical Character Recognition (OCR) system for printed Devnagari script is presented in this paper. Development of an OCR system for Devnagari is difficult because (i) there are about 350 basic, modified (matra) and compound character shapes in the script and (ii) the characters in a word are topologically connected. In our system the document image captured by a flatbed scanner is subjected to noise cleaning, skew correction, line segmentation, zone detection, and word and character segmentation using some conventional and some newly developed techniques. From zonal information and shape characteristics, the basic, modified and compound characters are separated for the convenience of classification. Modified and basic characters are recognized by a structural feature based tree classifier while compound characters are recognized by hybrid approach. At present, the system has an accuracy of about 96% at the character level.

5.1.36 Script Line Separation From Indian Multi-Script Documents

U. Pal and B. B. Chaudhuri, *IETE Journal of Research*

Abstract

In a multi-lingual country like India, a document page may contain more than one script form. Under the three-language formula, the document may be printed in English, Devnagari and one of the other Indian official languages. For Optical Character Recognition (OCR) of such a document page, it is necessary to separate these three script forms before feeding them to the OCRs of individual scripts. In this paper an automatic technique of separating the text lines is presented for almost all triplet of script forms. To do

so, the triplets are grouped into five classes according to their characteristics and shape based features have been employed to separate them without any expensive OCR-like algorithms. The proposed approaches are tested on many documents and the experimental results are presented. At present, the system has an overall accuracy of about 98.5%.

5.1.37 Script Line Separation From Indian Multi-Script Documents

U. Pal and B. B. Chaudhuri, *Int. Conf. on Document Analysis and Recognition, IEEE Computer Society Press, pp 406-409, 1999.*

Abstract

In a multi-lingual country like India, a document page may contain more than one script form. Under the three-language formula, the document may be printed in English, Devnagari and one of the other official Indian languages. For OCR of such a document page, it is necessary to separate these three script forms before feeding them to the OCRs of individual scripts. In this paper, an automatic technique of separating the text lines using script characteristics and shape based features is presented. At present, the system has an overall accuracy of about 98.5%.

5.1.38 Connected Script Synthesis By Character Concatenation - An Overlap And Weighted Average Formulation

V. Ramasubramanian and P.V.S. Rao, *Proceedings of SEARCC-88, New Delhi, Nov. 28-Dec.1, 1988, pp. 163-176.*

Abstract

In this paper we address the problem of synthesizing connected handwritten script from individual characters written isolation. Connected writing is viewed as a natural evolution from writing the characters in isolation, characterized by the use of continuous pen-down connecting movement from one character to the next. The problem is one of concatenation of individual character shapes to generate the connected script and consists in synthesizing a so called transition stroke from one character to the next. Particular emphasis is laid on recreating the context effect underlying the pen-down transition stroke and in preserving the continuity of motion and shape in the transition. Under this framework, we propose an approach based on the temporal description of the context effect as an overlap and weighted addition of appropriate segments of the temporally adjacent characters. The weighting (or blending) functions are designed and solved analytically such that the resulting transition curve satisfies a specified order of continuity at the transition points. We show that the weighting function is equivalent to a non-parametric Bezier curve with Bernstein polynomials as the basis function and a simple control point configuration which is a function only of the order of continuity sought. This equivalence result provides additional insight to the solution and a very simple way of generating the weighting function and the transition curve for any required order of continuity. The proposed approach is shown to generate connected script having a convincingly high degree of naturalness.

5.1.39 A Knowledge-Based Approach For Script Recognition Without Training

P.V.S. Rao, *IEEE Transactions on Pattern Analysis & Machine Intelligence (PAMI), Vol. 17, No. 12, pp. 1233-1239, December, 1995.*

Abstract

The approach is based on an empirical parametric model for the writing hand system. The parameters are so chosen and quantized as to retain only broad shape information ignoring writer-dependent and other variability. Concatenation of character prototypes generates archetypal reference words for recognition; training is unnecessary. Recognition scores exceed 90%.

5.1.40 Cursive Script : Recognition And Synthesis-Recent Trends

P.V.S. Rao, *Sectional Presidential Address at the 78th Science Congress, Indore, Jan.3-9, 1991.*

Abstract

It has been the usual practice in the Science Congress for Sectional students to choose as the topic for their Presidential Address, either an area in which they have themselves made contributions or an area which relates to the main theme chosen for that year.

Since this is the first Presidential Address to be delivered for the Computer Science Section, I thought it would be appropriate to choose a subject which is very much a current area for research. I felt also that it would be appropriate if the area has practical applications. In addition, it would be desirable to choose a subject that is not esoteric and highly theoretical, but one that is of consequence even to the non-specialist. It would then lend itself to ready appreciation without substantial effort to follow the jargon or the methodology. It would be an added attraction if the area relates to applications where computers appear to display capability which, if humans had it, would be called 'intelligence': i.e. the area of artificial intelligence. Accordingly, I have chosen the area of script recognition by computer: an area which is challenging, has numerous practical advantages, falls in the field of artificial intelligence,

and does not require much specialized knowledge to appreciate. It also happens to be an area in which I have been working in the recent past and to which I have made some non-trivial contributions. I shall therefore be talking about synthesis of cursive script words out of individual characters, characters of simpler elements and about recognition of characters and words by computers.

5.1.41 Script Recognition

P.V.S. Rao, Sadhana, *vol. 19 part 2, pp. 257-270, April, 1994.*

Abstract

This paper describes an approach for word-based on-line and off-line recognition of handwritten cursive script composed of English lower-case letters. The system uses simple and easily extractable features such as the direction of movement and curvature and the relative locations of regions where these suffer discontinuities.

Our approach was evolved based on our concept of 'shape vectors' introduced earlier. We visualize script characters as having shapes which are composed of comparatively straight segments alternating with regions of relatively high curvature. We derive the shape vectors from each script character essentially by identifying regions of least curvature and approximating these by straight lines. That these shape vectors carry adequate information about the identity of the character is established by showing that the original character can be faithfully reconstructed from the shape vectors.

We thus use slopes of the shape vectors and relative locations of points of maximum curvature (both highly quantized) as parameters for recognition. The system extracts parameters for individual characters from single specimens written in isolation and uses these to construct feature matrices for words in the vocabulary. These are used for matching with the feature matrices of test words during the recognition phase.

The advantage of the system is that it does not require elaborate training. Recognition scores are in the neighborhood of 94% for vocabulary sizes of 200 words. The approach has been extended for off-line information as well and performs quite well even in this case.

5.1.42 SHAPE VECTORS : An Efficient Parametric Representation For The Synthesis And Recognition Of Hand-Script Characters

P.V.S. Rao, Sadhana, *Vol. 18, Part 1, pp. 1-15, March 1993.*

Abstract

Earlier work by the author has established: (i) that cursive script can be synthesized out of individual characters by using polynomial merging functions which satisfy boundary conditions of continuity of the displacement functions x(t) and y(t) for each character and their first and second derivatives; and (ii) that the procedure lends itself to a Bezier curve based formulation. This was done since cursive writing avoids discontinuities (of shape) between individual characters as well as discontinuities in pen movement.

We show here that even individual characters can be synthesized out of more primitive elements by using the same merging functions. We choose directed straight lines which we call shape vectors as basic elements for this. Script characters generally have shapes which are essentially straight segments alternating with 'bends' or regions of relatively high curvature. For a character with n bends, we need only n+1 shape vectors. Thus each script character needs only three to seven shape vectors, depending on its complexity.

The "character generation" shape vectors are derived from the original character by means of a simple procedure that identifies comparatively straight regions in it. These are then approximated to straight line by linear regression and positioned to be tangential to the original curve. The synthesized version of this character is obtained by 'merging' or concatenating these vectors. The close fit between the original and re-synthesized characters demonstrates that the shape vectors adequately characterize their identities and shapes. Data reduction ratios in the range of 15 to 25 are thus possible. This method thus shows good promise as a possible basis for script character recognition, and a recognition scheme based on it has yielded an accuracy of 94% for a vocabulary size of 67 words.

5.1.43 Shape Vectors For On-Line And Of-Line Recognition Of Cursive Script Words

P.V.S. Rao, *First Intl. Conference on Document Analysis and Recognition (ICDAR 91), Saint-Malo, France, Sept. 30-Oct. 2, 1991, pp. 568-575.*

Abstract

This paper describes a novel approach for cursive script recognition and uses simple and easily extractable features such as the direction of movement and curvature and the relative locations of regions where these suffer discontinuities. This is based on earlier work which showed that (1) cursive script can be synthesized out of individual characters by using polynomial merging functions which satisfy boundary conditions of continuity of the displacement functions x(t) and y(t) for each character and their first and second derivatives, and that (2) even individual characters could be synthesized out of straight lines (which we call shape vectors). The recognition scheme does not require elaborate training and yields recognition scores around 94%.

5.1.44 Telugu Script Recognition - A Feature Based Approach

P.V.S. Rao and T.M. Ajitha, *International Conf. on Document Analysis and Recognition (ICDAR-95), Montreal, Canada, August 14-16, 1995.*

Abstract

Telugu characters can be visualized as being composed of circular segments of different radii. Recognition consists in segmenting the characters into the component elements and identifying them. We choose a feature set to preserve the canonical shapes while filtering out as noise the shape deviations encountered in real life. Hence, this approach does not require extensive training. Instead, 'Feature Vector' parameters for individual 'basic' characters are extracted from single specimens written in isolation. These are suitably combined to construct 'Feature Vectors' for compound characters for the lexicon. These are compared with similar 'Feature Vectors' extracted from the test samples to be recognized. Recognition scores ranged from 78 to 90% across different subjects, (when the best match alone is taken) and from 91 to 95% for a single subject.

5.1.45 A Complete Multi-Font OCR Systems For Printed Telugu Text

C.Vasantha Lakshmi, C.Patvardhan, *Language Engineering Conference, University of Hyderabad, India, Dec. 2002.*

Abstract

This work describes the design and development of a Telugu Optical Character Recognition system for Printed text (TOSP). Pre-processing tasks considered in this paper are: Conversion of gray scale image to a binary image, image rectification, skew detection and removal, segmentation of text into lines, words and basic symbols. Basic symbols are identified as the fundamental unit of segmentation in this paper which is recognized by the recognizer. The combinations of these basic symbols that together form characters and compound characters of Telugu are also determined to complete the recognition process. The special feature of TOSP is that it is designed to handle multiple sizes and multiple fonts. Further, the output produced by TOSP can directly be opened in any Indian language software that supports transliteration facility into Telugu script and edited. Several such soft wares are popular and available.

c) Speech Recognition

5.1.46 Relational Studies Between Phoneme And Grapheme Statistics In Current Bangla

B.B. Chaudhuri & U. Pal, *Journal of Acoustical Society of India, vol.-23, pp. 67-77, 1995.*

Abstract

This paper deals with the study of phonemic and graphemic character occurrences in words of Bangla (Bengali) language. The occurrence statistics of characters are presented for words collected from newspaper and popular magazines. Some of the computed grapheme, their position wise occurrences, percentage of compound character, word length versus frequency of occurrence, bi-gram etc. This study can be applied in Speech recognition, Speech analysis, Character recognition, Keyboard setting, Linguistics, Data communication, Cryptography and Error correction.

5.1.47 Animating Expressive Faces To Speak In Indian Languages

Tanveer A. Faruquie, Chalapathy Neti, Nitendra Rajput, L.Venkata Subramaniam, Ashish Verma, *NCC, Mumbai, Jan 26-27, 2002.*

Abstract

This paper describes a morphing based automated audio driven facial animation system. A novel scheme to implement a language independent system for audio-driven facial animation given a speech recognition system for just one language, in our case, English, is presented.. New viseme and expression combinations are synthesized to be able to generate animations with new facial expressions.

5.1.48 Spoken Word Recognition: Lexical Vs Sublexical

Amit Gupta and Sandeep M, Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003 .

Abstract

Spoken word has become a primary object of scientific inquiry with a focus on understanding how our speech perception capacities are used in segmenting and recognizing words in fluent speech. The present study investigated the nature of spoken word representation. Ten normal native Kannada speakers in the age range of 15-25yrs participated in the study. A word-spotting technique was used. Eighty Kannada words and non-words with 5 words and 5 non-words appearing twice were audio presented. The subjects were instructed to press the button when they heard the same word/ non-word for the second time and responses were audiorecorded which were then analyzed for the reaction time. The results of the present study indicated that words are spotted better than non-words supporting a lexical representation of words.

5.1.49 Data-guided Processing of Speech

Hynek H ermansky, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003 .*

Abstract

The paper introduces a new class of signal processing techniques that are trained on large amounts of speech data to extract a set of features for automatic recognition of speech. Such optimized signal processing techniques appear to be consistent with some important properties of human hearing.

5.1.50 Statistical Language Modeling Using Syntactically Enhanced Lsa

Dharmendra Kanejiya_,Arun Kumary, Surendra Prasad, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

Statistical language models using n-grams are inadequate to model long distance syntactic and semantic dependencies in a language. The syntactic dependencies can be modeled using some grammatical representation of text, and semantic dependencies can be captured using a technique called latent semantic analysis (LSA). However, to model both these dependencies simultaneously, we need a unified framework to represent them. Towards this direction, we present here a mathematical framework, called syntactically enhanced LSA (SELSA) that augments a word with the syntactic tag of its preceding word within LSA framework. This leads to a statistical language model that uses the preceding syntactic information along with the long distance semantic information to assign probabilities to words. Preliminary experiments on WSJ corpus show that SELSA reduces the bi-gram perplexity by 33:92% compared to 36:33% reduction by LSA, however it generates better probabilities for syntactic-semantically regular words than LSA.

5.1.51 Application Of Fourier Transformation In Digital Signal Processing For Speaker Identification

Sanghamitra Mohanty, Sanjaya Kumar dash, *National Conference on Mathematics and its Application , Feb 23-24, 2002.*

Abstract

Speaker identification is the process of determining the registered speaker providing a given utterance. Utterance of a speaker is termed as speech. Speech is a continuous acoustic signal which must be transformed into a sequence of discrete linguistic units. Speech signal in the form of speech wave form is converted to some type of parametric representation for further analysis and processing. One of the most important feature of speech signal i.e. frequency can be analysed by fast Fourier transformation. The spectrum is analyzed and standardized to maintain the standard database. This standard database is processed for text-to-speech system, speech recognition and speaker identification.

5.1.52 Recognition Of Voice Signals For Oriya Language Using Wavelet Neural Network

S. Mohanty & S.Bhattacharya, *STRANS 2001, Feb 15-17, IIT, Kanpur.*

Abstract

Speech recognition is both speech oriented and speaker oriented. Both have fuzzy effect on them, which adds to the hardness. During speech recognition, separation of the words and again redundancy of the voice responsible for the creation of words due to the vowels makes it difficult for analysis. We are trying with the wavelet neural network model to make an effective analysis for the recognition of speech signals. Here the main characteristics of speech/voice like frequency; intensity, accent and quality are analyzed for better output limiting the associated noise. Here we have used the concept of "wavelon" and "scalon". Using the control parameter of the network we have designed the wavelet network for two characters at the beginning and also we are extending for the rest of the character.

5.1.53 On Deriving A Phoneme Model For A New Language

Niloy Mukherjee, Nitendra Rajput, L. Venkata Subramaniam, Ashish Verma, *Proceedings: IEEE International Conference on Spoken Language Processing (ICSLP 2000), Beijing, China, Oct 16-20, 2000.*

Abstract

We present a method for building an initial phoneme model for training an HMM in a new language using an already trained recognition system in a base language. HMM based phoneme recognition systems are used to model the phonemes in most large vocabulary speech recognition tasks. Mappings between the phonetic spaces of the two languages are generated and are used to populate the phonetic space of the new language. The best possible alignment of the new language data is obtained and initial phone models are built on this labeled data. A classification experiment is performed in the new language to illustrate the goodness of initial phone models.

Experiments are carried out with Hindi as the new language using an English language recognition system to derive the initial phone models for Hindi language.

5.1.54 Dictionary Supported Generation Of English Text From Pitman Shorthand Scripted Phonetic Text

P. Nagabhushan, Basavaraj S. Anami and D. S. Guru, *Language Engineering Conference, University of Hyderabad, India, Dec. 2002.*

Abstract

The Pitman Shorthand Language (PSL) is a recording medium practiced in all organizations, where English is a medium of transaction. It has the practical advantage of high speed recording, more than 180 words per minute, because of which it is appreciably received all over. This recording medium has its continued existence in spite of considerable developments in speech processing systems, which are not universally established yet. In order to exploit vast transcribing potential of PSL a new area of research on automation of PSL processing, is conceived. This paper describes the substitution of equivalent English words for the phonetic compositions of the transcribed words, in the process of automatic generation of English text from PSL document. The transcription is achieved by making use of two new types of dictionaries specifically developed and implemented for this purpose, one of them being a phonetic dictionary wherein the words are sequenced in phonetic order and the other being an extended conventional dictionary wherein the words are appended with additional details such as domain of their use, forms of verbs and so on. The proposed approach is tested with limited words in both the types of dictionaries and is found to perform satisfactorily. However, the scope exists for addition of new words into these dictionaries.

5.1.55 A Pairwise Multiple Codebook Approach To Implicit Language Identification

T.Nagarajan and Hema A. Murthy, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003 .*

Abstract

Automatic spoken language identification is the task of identifying the language from a short duration of a speech signal. One of the important language identification cues is the deference's in phoneme frequencies among different languages. Considering this, we develop a pair-wise multiple codebook approach to language identification. This system is compared with the traditional single code book per language system. Traditional VQ based language models are generally preferred since they do not explicitly require language models. The evaluation with Oregon Graduate Institute Multi- Language Telephone Speech Corpus shows that the multiple codebook system improves the performance by almost 7%.

5.1.56 Product-hmm A Novel Class Of Hmms For Subsequence Modelling

Mukundhn Nagarajan And T.V.Sreenivas, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003 .*

Abstract

This paper presents a novel kind of HMM, called Product HMM that can be used for subsequence modeling. A sub vector is formed by selecting particular components from the original vector. A sequence of such sub-vectors forms a sub-sequence. This paper considers the case of modeling a vector sequence in terms of its two sub-sequences. In the present framework of HMM, the architecture of the HMM is fixed for a class of vector sequences. The freedom of selecting a suitable architecture for each of the sub-sequences (of the class of vector sequences considered earlier) is not possible. Product-HMM offers this flexibility of selecting a separate architecture for each of the subsequences. Number of states and structure of the transition matrix (which decides whether the HMM is a left-to-right one or an ergodic one) constitute the architecture of a HMM. So, Product HMM offers the freedom to choose different number of states for each of the sub-sequences and to choose independently whether each of them is to be modeled by a left-to-right or an ergodic HMM. It is shown that modeling using Product HMM is better than modeling the two sub-sequences using two independent HMMs. This way of joint training of a HMM from two streams of sequential data has not been tried before. Product-HMM is an integrated statistical model which provides a way of integrating different HMMs that model the subsequences of a vector sequence. The possibility of having optimal HMM architectures for the sub-sequences results in utilizing the training data better and hence in better estimates of model parameters.

5.1.57 Synthesis Based Recognition of Continuous Speech

K.K. Paliwal and P.V.S. Rao, *Journ. Of Acoust. Soc., America, Vol. 71, No. 4, pp. 1016-1024,April 1982.*

Abstract

An acoustic phonemic recognition system for continuous speech is presented. The system utilized both steady-state and transition segments of the speech signal to achieve recognition. The information contained in formant transitions is utilized by the system by using a synthesis-based recognition approach. It is shown that this improves the performance of the system considerably. Recognition of continuous speech is accomplished here in three stages: segmentation, steady-state recognition, and synthesis-based recognition. The system has been tried out on 40 test utterances, each 3-4 s in duration, spoken by a single male speaker and the following results are obtained: 5.4% missed segment error, 8.3% extra segment error, 52.3% correct recognition using only steady-state segments, and 62.0% correct recognition using both steady-state and transition segments. PACS numbers: 43.70 Sc

5.1.58 Adapting Phonetic Decision Trees Between Languages For Continuous Speech Recognition

Nitendra Rajput, L. Venkata Subramaniam, Ashish Verma, *Proceedings: IEEE International Conference on Spoken Language Processing (ICSLP 2000), Beijing, China, Oct 16-20, 2000.*

Abstract

In a continuous speech recognition system it is important to model the context dependent variations in the pronunciations of phones. In this work we have attempted to build decision trees for modeling phonetic context-dependency in Hindi. The approach followed is to modify a decision tree built to model context-dependency in American English. The reason the decision trees turn out to be different are that the English and Hindi phoneme sets are not identical. Then even for identical phonemes, the context-dependency is different for the two languages. Linguistic-Phonetic knowledge of Hindi is used to modify the English phone set. Since the Hindi phone set being used is derived from the English phone set, the adaptation of the English tree to Hindi follows naturally. Though here the adaptation is from English to Hindi, the method may be applicable for adapting between any two languages. The decision tree is built using either Hindi data or English data labeled with the correct Hindi contexts. This procedure is discussed and the limitations of both the methods are described.

5.1.59 A Large Vocabulary Continuous Speech Recognition System For Hindi

Nitendra Rajput, Ashish Verma, Chalapathy Neti, *NCC, Mumbai, Jan 26-27, 2002.*

Abstract

Abstract: In this paper we present a Hindi Speech Recognition system which has been trained on 40 hours of audio data and a has a trigram language model that is trained with 3 million words. For a vocabulary size of 65000 words, the system gives a word accuracy of 75% to 95%.

5.1.60 Language Identification Using Parallel Phone Recognition

V. Ramasubramanian A. K. V. Sai Jayram T. V. Sreenivas, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

We study some of the unexplored issues in the parallel phone recognition (PPR) system for automatic language identification (LID). We consider three types of scores for LID, namely, the acoustic score, language model score, and the joint acoustic language score. Using each of these scores we formulate three types of classifiers for performing LID: maximum likelihood classifier (MLC), Gaussian classifier (GC) and K nearest neighbor classifier (KNNC) and compare their performances. We examine the problem of bias in the PPR scores which affects LID performance and interpret the bias and bias removal methods which improve the classification accuracy of MLC. Among all the different combinations of scoring methods and classifiers, it is found that MLC with bias removal performs best for either acoustic or language model score alone; this is closely followed by GC with the joint acoustic language score.

5.1.61 Design And Implementation Of Digital Speech Security System

K. T. V. Reddy, V. Prasad and N. Padmakar, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003 .*

Abstract

In digital speech security system (DSSS), the analog signal is converted into pulse code modulated (PCM) digital form using an analog-to-digital converter and encrypted before transmission by a special technique. The desired party can decipher the message by treating the received encrypted data with the same technique. Thus conversation can be carried out without interception. The digital signal is combined directly with the output from a pseudorandom number generator (noise generator) to obtain an encrypted speech before transmission. Such a system has the property that for the interceptor, the received message appears like noise and thus prevents him from eavesdropping. However, the desired party can decipher the message by once again mixing the received enciphered message with the local replica of the pseudorandom noise (PN) available with him.

5.1.62 Blank Slate Language Processor For Speech Recognition

P.V.S. Rao and Nandini Bondale, *ICSLP'92, Banff, Canada, October, 1992.*

Abstract

We describe here a novel 'language processor' (LP) which uses syntactic and semantic properties of Indian language. Speech recognition system for railway reservation inquiry task. The acoustic level recognizer provides several alternatives for each word with associated 'confidence levels'. Using these, the LP has to generate a 'most likely' sentence consistent with the constrains imposed by syntax and semantics. The LP requires no a priori knowledge of the syntax or training phase. During the recognition phase, it uses this knowledge at the sentence level to correct for word recognition errors made by the acoustic level recognizer.

5.1.63 Is Rule Based Acoustic Phonetic Speech Recognition A Dead End?

P.V.S. Rao, *The Second Symposium on Advanced Man-Machine Interface through Spoken Language, Nov. 19-22, 1988, Hawaii, USA.*

Abstract

Speech recognition using sub word units and Hidden Markov models has become very popular in the recent past in view of their versatility and good performance. Acoustic-phonetic rule based recognition, which has been the main stay for many years, appears no longer to be so attractive. In this paper, we attempt a study of all three approaches and their implications to determine whether the acoustic-phonetic approach has a future at all. We conclude on the bias of our analysis that it holds a potential for the future. We also visualize the prospect of an integrated approach which combines the strategies of all three.

5.1.64 Language Processing For Speech Recognition

P.V.S. Rao and Nandini Bondale, *– An Invited Paper, Computer Science & Informatics, Journal of Computer Society of India, Vol. 24 No. 4, pp. 1-13, December, 1994.*

Abstract

In this paper, we are primarily concerned with the problem of how language related information can be used to improve the performance of speech recognition systems. Our aim is to give a brief amount of the various techniques used in this context and to present some of the work being done in our laboratories in this area. Use of 'n-grams' is a statistical approach to reduce perplexity (or the uncertainty factor in the recognition of entities such as words – in speech recognition). Their use, even in isolation yields significant improvements in recognition scores. However, impractically large sizes of corpora are needed to get these statistics. We investigate two alternative approaches in this context: using what we call a 'personalized grammar; and to use word category n-grams instead of word n-grams. The latter approach captures the semantic as well as syntactic compatibility constraints in language. To this end, we try out an automatic method of grouping words (say nouns) into equivalence classes based on criteria relating to their compatibility with other words or word groups (say adjectives or verbs). This technique was applied for a text of Ramanyana as well as for a more modern vocabulary related to the field of computers and conferences. In both cases, the nouns get appropriately grouped and the machine classification so achieved is in conformity with intuitive groupings.

Further, our aim is to evolve a scheme that captures the essential features of the structure of language (syntax and semantics) in a manner that readily lends itself to use in speech recognition. It would be difficult to do this by formulating explicit rules of grammar, particularly for restricted (and task related) discourses of the type that are used in speech recognition applications. It would also be desirable to minimize subjective judgement in these processes. Accordingly, we describe here an approach which derives the structure from the corpus itself. This has been used to optimize the performance of a Hindi speech recognition system for railway reservation inquiry task. Use of this system increases the accuracy of recognition very significantly. The LP can also be used as a tool in research relating to the analysis of child language corpora to derive the models of grammar used by children in various stages of language.

5.1.65 On Stop Consonants

P.V.S. Rao, *Speech Comm. Seminar, Stockholm, 1974, pp 95-102.*

Abstract

Noise burst frequencies and, more importantly, vowel formant transitions are characteristic cues for the perception of stop consonants. While formant transitions are to a certain extent influenced by context effects due to non-immediate neighbors, it has been shown in this paper that these effects are non important for human perception; the human listener looks at the speech signal only two phonemes at a time. He is known to discriminate between two stimuli separated by a constant distance in an acoustic measurement space far better when they are separated by a phoneme boundary than otherwise. A cross-linguistic experiment reported here indicates that this is an acquired rather than innate, ability. Existing methods for speech recognition by machine are not very satisfactory for discriminating between stop consonants; this involves parameters such as formant loci and initial or terminal frequencies of formant transitions which are not easy to determine. A principal component analysis method is presented which enables satisfactory identification of stop consonants based on the instantaneous values of formant frequencies and slopes at any given point within the formant transition. Using this method, it has been shown that it is possible to rank the parameters in the order of their importance for stop consonant identification.

5.1.66 Speech Research

P.V.S. Rao, *at the Tata Institute of Fundamental Research, Transactions of the Committee on Speech Research, The Acoustical Society of Japan, University of Tokyo, March 23, 1982 S81-106 (1982-10).*

Abstract

The speech and digital systems group of Tata Institute of Fundamental Research has been actively involved with work in a number of topics connected with speech: speech synthesis, recognition, perceptions and signal processing. It implemented a synthesis scheme which achieved synthesis by rule, the rules being specified for classes of phonemes rather than for individual phonemes, and worked on speech perception and the nature of the relation between articulatory phoneme boundaries and perceptual phoneme boundaries for stop consonants. This work indicated that the heightened capability of discrimination for stimuli lying on either side of the perceptual boundaries, being language specific, was an "acquired" rather than an inherent one. Recent work included extending the recognition scheme for continuous speech and a number of problems in speech analysis: linear prediction analysis-synthesis, formant estimation, maximum entropy method spectral analysis and pitch estimation techniques.

5.1.67 VOICE : An Integrated Speech Recognition Synthesis System For The Hindi Language

P.V.S. Rao, *Speech Communication, 13, pp. 197-205, May, 1993.*

Abstract

A Voice Oriented Interactive Computing Environment (VOICE) has been implemented in the Hindi language. The system provides an interactive facility for visual and voice feedback. The 200 isolated word recognition system is designed around a railway reservation inquiry task and uses acoustic-phonetic segments as the basic units of recognition. Frame level classification into broad acoustic-phonetic categories is accomplished by a maximum likelihood vectors by use of explicit duration semi (Hidden) Markov Models. A more detailed classification of a few categories (vowels, voice bar and nasals in the first instance) is performed by neural nets. String matching using dynamic programming accomplishes lexical access, or conversion of the phonetic category symbol strings into words. Distributed processing of the word recognition task enables recognition at four times real time. A language processor disambiguates between multiple choices given by the recognizer for each word and even corrects some acoustic level recognition errors. This, the first system working in any Indian language, gives a recognition performance of 85% at the word level. For comparison, a purely HMM based word level recognizer has also been implemented. The performance is expected to improve further as there is still substantial scope for refinement.

5.1.68 A Comparison of Public Domain Software Tools for Speech Recognition

Samudravijaya K and Maria Barot– , *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003 .*

Abstract

HTK and Sphinx are two freely downloadable soft ware packages with the capability of implementing a large vocabulary, speaker independent, continuous speech recognition system in any language. While HTK has been in use by various groups for about a decade, and has gone through the refinement cycles necessary for a commercial software, Sphinx was released about an year ago and is still undergoing development in a university environment. However, due to certain advanced features and the license for unrestricted use, Sphinx appears to be more attractive. These two software packages have been compared by implementing a Hindi speech recognition system. Although recognition accuracies of the two systems are comparable, we observe that the acoustic modeling of Sphinx is superior.

5.1.69 Acoustic Modeling of Subword Units Using Support Vector Machines

Chandra Sekhar, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

This paper addresses the issues in recognition of subword units of speech using support vector machines. Discriminative training and good generalization capability of support vector machines are useful in developing acoustic models for subword units when the confusability among the units is high. We compare the performance of support vector machine based systems with that of hidden Markov models in recognition of subword units. We demonstrate the better performance of support vector machines in recognition of monophone units in a large corpus of Japanese speech and recognition of consonant-vowel units in a broadcast news corpus of an Indian language.

5.1.70 V-Mail: Voice Email Communication over the Phone

Sheetal S. Sethi, K S R Anjaneyulu, Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003 .

Abstract

The main objective of this paper is to give an insight into the V-Mail project and how it tries to use speech recognition for providing a voice e-mail service to users. We first describe the motivation behind the project and the audience targeted to use the system.

We then outline the applications for which the system could be put to use, and its architecture and design. We finally discuss the speech recognition issues, the current and the future plans for the project.

5.1.71 A Study Into Front-end Signal Processing For Automatic Speech Recognition

Rohit Sinha and S. Umesh, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003 .*

Abstract

In our recently reported work, we have observed some difference in recognition performance using our proposed method of feature computation when compared to features computed using the traditional mel-filterbank analysis. In the alternate method of feature computation, we use a spectral smoothing procedure which is very similar to weighted overlapped segment averaging (WOSA) method of spectral estimation. In this paper, we study the signal processing of the above mentioned feature computation methods, and point out to the differences between the two methods, and the effect of these differences on the recognition performance.

5.1.72 A Software System For Speech Recognition

R.B. Thosar And P.V.S. Rao, *7^{th} Internatl. Conf. Acoust. Budapest, 1971, Hungary, Pp 9-12.*

Abstract

Speech recognition by machine attempts to recover, from a continuous acoustic signal, the underlying message, coded as a string of phonemes. It is impossible to isolate parts of the acoustic signal which have a one-to-one correspondence with individual phonemes; in addition to their individual acoustic characteristics, adjacent phonemes show considerable mutual interactions. Such interactions play an important role in human speech perception and hence must be made use of, to the extent possible, in machine recognition of speech.

The transitions in vowel formants, generated in conduction with consonants, are some of the most important and extensively studied context effects. The formant transitions characterize the consonant and have primary significant for the perception of stops. Systematic variations in vowel duration and steady state formant positions of laterals have also been analyzed.

This paper presents an approach to speech recognition utilizing such contextual effects.

5.1.73 An Approach Towards A Synthesis Based Speech Recognition System

R.B. Thosar and P.V.S. Rao, *IEEE Trans. Acoust. Speech and Signal Processing, ASSP-24, April 1976, pp 194-196.*

Abstract

Preliminary results are presented of experiments with a recognition scheme intended for continuous speech. The scheme utilizes information about interphoneme contextual effects contained in formant transitions and employs internal trial synthesis and feedback comparison as a means for recognition. The aim is to achieve minimal sensitivity to the appreciable variability which occurs in the speech signal, even for utterances of a single speaker. While the approach outlined here is quite general, it has initially been tried out on vowel-stop-vowel utterances. Recognition scores obtained are encouraging and demonstrate the viability of the approach.

d) Speech Synthesis

5.1.74 Synthesizing Hindi Speech Using Klsyn and Hlsyn for Natural Sounding

S.S.Agrawal, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003*

Abstract

High quality synthesis of different Spoken Languages has been a challenge to speech scientists and engineers. This has become more important where the task is to develop unlimited text to speech synthesis system. Several techniques based on simulation of articulary models, electrical analogues (formant synthesizers) and concatenation of natural speech waveform samples from a large corpus have been developed and used with varying success. However, extensive research is still going on to produce natural sounding multilingual synthetic speech with desired accents, mood, situation and quality etc.

The present paper, describes the possibility of using two important synthesizers i.e. the Klatt Synthesizer (KLSYN) and the High Level parametric Synthesizer (HL- SYN) for synthesizing high quality Hindi Speech and other Indian Spoken languages such as Bengali. Both the Synthesizers implemented on a PC have been used to synthesize the syllables containing all Hindi consonants containing special features such as aspiration and retroflexion etc. and the words containing geminates and clusters and continuous sentences by copy synthesis. In a similar manner Bengali speech sounds were also synthesized. The parametric data base of Hindi Syllables was also used to synthesize Bengali Speech. A careful control of the acoustic and pseudo articulatory parameters produced quite satisfactory quality synthetic speech. Superimposition of prosodic parameters extracted separately from original speech of Hindi and Bangla improved the naturalness of synthesized speech. It has been demonstrated that both KLSYN and HLSYN can be used to produce naturally sounding Hindi Speech and possibly other Indian Spoken languages also.

5.1.75 Some Important Aspects Of Bengali Speech Synthesis System

Asok Bandyopadhyay, *Speech and Signal Processing Group, ER&DCI, Calcutta.*

Abstract

This paper deals with some important issues of Text-to-Speech Synthesis System in Bengali language. The basic building blocks of a synthesizer and their implementation strategies have been discussed here. The approach for text analysis, which is relevant to the Bengali language, has also been included. Author(s) tried to make clear the importance of interpretation of phonological information in Bengali.

In the context of realization of Text-to-Speech Synthesis System in Bengali the concatenation of signal segments is done using ESOLA (Epoch Synchronization and Overlap Add) method are applied to broad windows to help in Spectral smoothing of the transition between the segments.

The concatenation module consists of joining two partnemes (part of a phoneme) using certain rules. Signal elements of the dictionary are terminated at both ends at positive going zero crossing to avoid phase mismatch. The constant pitch allows a large overlap of the signals.

The synthesizer developed has been tested on Bengali sentences. The signal dictionary has been extracted from a corpus recorded by Bengali speaker, trying hard to keep a constant pitch during the pronunciation to facilitate the initial pitch modification. The synthesized sentence is fairly clear with good degree of phonetic naturalness.

5.1.76 Effects Of Pitch Contours Stylization And Time Scale Modification On Natural Speech Synthesis

Asok Bandyopadhyay, Shyamal Kr. Das Mandal and Barnali Pal, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

This paper describes the method of generation of intonated speech for natural speech synthesis using prosody generation model. The effect of pitch modification through pitch contour stylization for parameter extraction and time scale modification for it's implementation has been mentioned. An approach for close-copy syllabic stylization has been described. In the latter part, algorithm for implementation of time scale modification with necessary approximation for sinusoidal signal has been mentioned. Experimental results of applying the technique for pitch modification on Bengali sentence have been shown. The output shows satisfactory performance of sound quality after necessary pitch modification to make synthetic speech natural.

5.1.77 Utterance Rules For Bangla Words And Their Computer Implementation

Dash, N.S. and B.B. Chaudhuri *(1988) Published in the Proceedings of the International Conference of Computational Linguistics, Speech and Document Processing (ICCLSDP'98), C: 51-58, 1998.*

Abstract

This paper deals with the utterance rules of Bangla words and their computer implementation. It is observed that though Bangla script is mostly phonological in nature, the utterance of some graphemes of a word is not always identical with their utterance in isolation. Various scholars tried to find utterance rules that are limited mainly to word-initial position. It is important and useful to find rules for word-final as well as word-middle positions. This report is an attempt to fill up this gap. Our work is motivated by designing a computer based Bangla speech synthesizer where the synthesized speech would be more lively if these utterance rules are implemented. To find new rules, we have used a sample corpus of 5,00,000 words and a dictionary of about 60,000 words. We have tried to describe the utterance rules according to the position of graphemes in the word. The rules discussed by previous scholars are compiled and many new ones discovered by us are described. Finally, the algorithm for implementing them in computer synthesis is considered.

5.1.78 Prosody, Information, and Modeling—With Emphasis on Tonal Features of Speech

Hiroya Fujisaki, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

Starting from the author's view on the process of information manifestation in the tonal features of speech, this paper emphasizes the importance of objective and quantitative modeling in the study of these features. It then describes a model for the process of fundamental frequency control of speech that has been originally proposed and established for Japanese, and explains the physiological and physical evidences on which the model is based. Application of the model for generation of F_0 contours of languages other than Japanese is then described, indicating how the original model can be modified and extended to cover those features that are not found in Japanese. The underlying mechanisms responsible for production of these features are also discussed.

5.1.79 Improving Quality of Speech Synthesis In Indian Languages

P. K. Lehana and P. C. Pandey, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

Harmonic plus noise model (HNM) which divides the speech signal in two sub bands: harmonic and noise, is implemented with the objective of studying its capabilities for improving the quality of speech synthesis in Indian languages. Investigations show that HNM is capable of synthesizing all vowels and syllables with good quality. All the syllables are intelligible if synthesized using only harmonic part except /*a a*/ and /*asa*/. This fact can reduce the size of the database. For pitch synchronous analysis and synthesis glottal closure instants (GCIs) should be accurately calculated. The quality of synthesized speech improves if these instants are obtained from the glottal signal (output of an impedance glottograph) instead of these being obtained by processing the speech signal. Further the noise part is synthesized pitch synchronously for voiced frames. A database of HNM parameters for VCV syllables is developed for Indian languages. The number of parameters for each frame is comparable to that of formant synthesizer but the quality of synthesized speech is much better.

5.1.80 Tamil Voice Engine

Madhan Karky V,Thiyagarajan R,Manoj Annadurai,*'Chennai Kavigal'* Dr.Ranjani Parthasarathi, Dr.T.V.Geetha, *Tamil Inayam, Malaysia 2001.*

Abstract

This paper presents the design and development of a Voice Engine for Tamil language. A Tamil voice engine is expected to pave way for developing creative applications to enable users to hear Tamil content in voice form. The key concept behind this voice engine is Text-To-Speech conversion. This conversion uses the method of concatenating the syllables to generate the required words. The major part of this work is thus the identification of syllables in the language. Around four thousand syllables have been identified. Tamil text in TAM encoding is given as input to the voice engine. The engine transforms the text to speech by applying various layers of processing over the text. The layers are Normalization, Transliteration, Syllable Splitting, Buffering and Concatenation. First, the incoming text is processed for normalization where commonly used abbreviations are expanded and numbers are converted to corresponding textual form. The normalized text then goes to the transliteration layer where the text is properly transliterated to an English alphabet or Roman alphabet. This layer is meant to take care of homographic ambiguities in the Tamil language. A mapping database has been built to ensure uniqueness in the transliteration process. The third layer is applying the transliterated text to a syllable split automata. The automata designed for splitting a word to its corresponding syllables, will take a word as input and give out an array of syllables. The final phase is buffering and concatenation, where the syllables of each word are buffered, then concatenated while writing to the soundcard. To demonstrate the power of the engine, a few applications like a Talking browser, mail client, chat and notepad have been developed. The Talking browser parses for Tamil contents on the page and submits the Tamil text to the engine, which converts the Tamil text to its corresponding voice. A Talking chat application sends text via the network, and converts the text to voice on the local machine. A Talking Dictionary, Talking Notepad and Talking Mail client are other applications that have been developed. The concept of such an engine is to enable application developers to think of more user-friendly applications. The aim has been to give a tool for programmers, using which they can build any applications of their choice.

5.1.81 A Novel Approach For Text-To-Speech System For Indian Languages

S.Mohanty, S.Bhattacharya, D.Acharya, A.K.Senapati, Indian Science Congress, 2003.

Abstract

Indian Languages are phonetic by nature. In the Text-To-Speech System for Indian language (Oriya) the algorithm developed by us involves analysis of a sentence (combination of words), words(combination of characters) and characters involving combination of pure consonant and vowel technique. Speech signal is a waveform having in it the information regarding frequency, amplitude and timber bearing in it the acoustic value for the phonemes and diaphones. The merging of such wave files as per requirement to generate the modified consonants influenced by *matras*, *phalas* and *yuktaksharas* (conjuncts) generate the speech from a text.

5.1.82 DSP Approach To Speech Synthesis For Indian Languages: Oriya- A Case Study

S. Mohanty, S. Bhattacharya, D. Acharya, A. K. Senapati, *submitted to Workshop on spoken languages processing , TIFR, Mumbai.*

Abstract

Indian Languages are phonetic by nature. During Oriya Speech Synthesis system utterance of consonant by three different speakers are taken into consideration for an experimental standardization. Initial unintentional

noise is first cleared from the wave file by passing it through filters of specific requirement. Speech signals like other signals being in wave form have information regarding frequency, amplitude, pitch and timber bearing in it the acoustic value for the phonemes and diaphones. The merging of such wave files as per requirement generate modified consonants influenced by matras, phalas and yuktaksharas in the form of desired speech output. To achieve this goal we have analysed the frequency and pitch of a particular speaker using Short Time Fourier Transform (STFT). We computed the energy associated with the signal by power Spectral Density function and subsequently digital filters are designed to utter desired output of a particular consonant thus a word and then sentence uttered by the speakers.

5.1.83 A Programming System For Studies In Speech Synthesis

P.V.S. Rao and R.B. Thosar, *IEEE Trans. Acoust., Speech and Signal Processing, ASSP-22 No. 3, June 1974, pp. 217-225.*

Abstract

This paper describes a speech synthesis system which is particularly suitable for experimental investigations. The synthesis is accomplished in two stages. The concatenation stage generates a schematized spectrographic representation corresponding to the symbolic input. The second stage consists in generating the corresponding acoustic signal. The steady state characterization of each phoneme is supplied as data. Independent concatenation procedures incorporate context dependent effects such as format transitions, changes in the normal duration of vowels, etc. The parameter values for these procedures are obtained by a set of rules. Applicability of a rule is determined by attributes assigned to the phonemes.

The phonemes are divided into classes and subclasses by the attribute assignment. The attribute STOP, for instance, defines the class of all stop consonants and BILABIAL STOP would define the set /p, b, m/. Thus a rule specifies a parameter value when a sub class of phonemes occur, in the context of another subclass. Such a formulation considerably reduces the number of rules. The classification as well as the rules are supplied as data to the system, giving it considerable flexibility.

The spectrographic output of the concatenation stage is used to actuate a simulated series terminal analog synthesizer. Rudimentary prosodies are incorporated which modify a monotonous pitch contour with stress makers and interrogative or declarative termination of a sentence.

5.1.84 BSLP Based Language Grammars For Child Speech

P.V.S. Rao and Nandini Bondale, *ICSLP-94, International conference on Spoken Language Processing, Yokohama, Japan, September 18-22, 1994.*

Abstract

We report in this paper our investigations aimed at evolving an approach to develop a grammar model for the speech of a child acquiring English as the first language. We use the occurrence of words in similar syntacto-semantic environment as the criterion for classifying words into equivalence classes. Consequently, words in the same class are interchangeable in a phrase or larger context. We identify phrases on the basis of n-gram frequencies of 'class exemplars' in the corpus and their meaningfulness. We group them into a small number of interchangeable phrase types. It turns out that the corpus of 723 sentences consists of 161 distinct phrase type sequence patterns and that of these, just 70 patterns have a frequency more than one and account for as many as 632 sentences in the original corpus. We use an algorithmic procedure to build a 'State Transition Network' which accounts for the sentence patterns. The STN consists of 11 states and generates most of the sentences in the corpus. This indicates that the approach is effective for modeling the grammar of the child.

5.1.85 Computer Recognition Of Spoken Hindi Digits

P.V.S. Rao and K. Samudravijaya, *Proceedings of International Workshop on Computer Processing of Asian Languages, 26-28 Sept. 1989, pp. 299-304, Bangkok.*

Abstract

A computer system for recognizing spoken Hindi digits is described. The system uses acoustic-phonetic features to identify the unknown word. It follows the approach used by expert spectrogram readers to recognize the utterance. The list of candidate words is pruned based on robust features. The cues offered by other features are used in a judicious manner to choose the best candidate word.

5.1.86 Speech Synthesis Using A Structured Phoneme Classification Scheme

P.V.S. Rao and R.B. Thosar, *6th Internatl. Conf. Acoust. Tokyo, 1968, pp 207-210.*

Abstract

This paper outlines a synthesis scheme which has been evolved using a flexible programming package called SPEECH (for Speech Playback Evaluation and Experimental Concatenation Handler), developed

specially for the purpose. The main features of this approach are: (1) phonemes are divided into equivalence classes and subclasses to facilitate formulation and application of context dependency rules; (2) these rules are explicitly provided for as separate sub-routines. Further synthesis, even up to the level of the final acoustic signal, is accomplished within the computers. Only the final digital to analog conversion is done by hardware, thus minimizing equipment imposed limitations to quality and flexibility. The input message string can be typed in from the console typewriter of the computer.

5.1.87 Hindi Speech Database

Samudravijaya K, Rao P.V.S, and Agrawal, *S.S, International Conference on Spoken Language Processing, October 2000, Beijing*

Abstract

A general purpose, multi-speaker, continuous speech database was designed and developed for Hindi language. The sentences have been segmented and labeled in terms of sub-phonetic units which take into account the special acoustic-phonetic features of Hindi. A preliminary study of the durations of the Hindi vowels indicate that native Hindi speakers are better able (than non-native Hindi speakers) to utilize the durational cues to distinguish between members of vowel pairs.

5.1.88 Architecture Of An Application Specific Instruction Set Processor For Parametric Speech Synthesis

R. Saini y , S. Srivastava, A. S. Mandal, Sudhir Kumar, R. Singh, A. Karmakar, C. Shekhar *Scientist, y Trainee, Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

This paper analyses the parametric speech synthesizer by D. H. Klatt [1, 2] from the point of view of designing an Application Specie Instruction Set Processor (ASIP) chip for parametric speech synthesis. By analyzing Klatt's code, the frequency of different computational operations in the code is estimated and constraints on speeds in performing these operations are derived. Next, the definition of a suitable instruction set for the ASIP and the hardware architecture for its implementation are proposed and analyzed. The architecture is verified using VHDL modeling and simulation. The strategy of implementation of the instruction set on the proposed hardware architecture is discussed and estimates of equivalent gate counts and RAM blocks needed for FPGA realization are given.

5.1.89 Pronunciation Rules for Indian English Text-to-speech System

Aniruddha Sen, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

Text-to-speech synthesis in Indian English is useful for delivering messages stored in computers and web to the Indian users unfamiliar with standard English accent. Such work is going on at TIFR and the paper reports the salient features of the front-end language processor that generates pronunciation plus stress information. The important components of the language processor are the parser to categorize words, an Indian English phonetic dictionary, morphological analyzer, letter-to-sound rules, phonological rules, prosody rules and Indian name detector. The relevant rules are formulated with the aid of a large CMU pronunciation dictionary and a language tool GENEX, developed in-house, that can generate a sub-dictionary following a set of specified constraints. The paper outlines the rule formulation procedure and provides examples of various types of rules. A few important morphological rules and letter-to-sound rules are described in detail.

5.1.90 Process & Problems Involved In The Creation Of Diaphone Database For Malayalam Text To Speech System

Jose Stephen, K. G. Sulochana, R. Ravindra Kumar, *National Workshop on Application of Language Technology in Indian Languages, to be held in Hyderabad, during March 6-8, 2003*

Abstract

A Text-To-Speech (TTS) synthesizer is a computer-based system that reads out any Malayalam text input to it. The primary goal of a TTS system is to deliver the message contained in the text in the best possible way in order to ensure that the listener grasps the general sense. This paper discusses the process and problems involved in the creation of diaphone database for Malayalam Text to Speech system. We prepared the Malayalam diaphone database for the synthesis of speech using MBROLA speech engine. The various stages in creating the diaphone database like creating a text corpus, recording the corpus, segmenting the speech corpus into diaphones etc. are discussed in detail along with the short description of phase marking and smoothening of diaphones. Major set of phonemes and allophones present in Malayalam language are also listed out along with their pronunciation. The paper also gives a short description of different approaches

for Speech Synthesis such as *synthesis-by-rule* and *synthesis-by-concatenation*. The advantages and disadvantages of these approaches for speech synthesis are also described in this paper. The major problems identified in the preparation of the database and the developments of text to speech system are also enlisted.

5.1.91 Comparative Analysis of Hindi Retroflex and Dental Cv Syllables and Their Synthesis

Rajesh Verma and Puneet Chawla, Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.

Abstract

This paper describes in detail the analysis results of the Hindi Retroflex consonants /t./, /t.^{h}/, /d./ & /d.^{h}/ and the Dental consonants /t/, /t^{h}/, /d/ and /d^{h}/ analyzed by using PC based Sensimetrics Speech Station Software. These sounds were analyzed in five long vowel contexts /a/, /i/, /u/, /e/ and /o/ for a very accurate description of their acoustic characteristics/ features and the differences between the corresponding cognate sounds in the two classes. Various parameters like duration of closure/voice bar, duration of burst, voice onset time, duration of aspiration, rate of second formant transition and burst frequencies and amplitudes have been studied in details.

The analyzed data was further used to generate the synthetic CV syllables using a cascade/ parallel formant synthesizer simulated on a PC. For the synthesis purpose, the source and vocal tract parameters of the synthesizer configuration were selected very carefully. Special attention was paid to the parameters like formant frequencies and their relative amplitudes, which play an important role in making distinction between cognate sounds like /t/ and /t./. The overall burst amplitude also plays a crucial role to make clear distinction in dental and retroflex cognate sounds. The parametric doc files were modified iteratively, until a satisfactory quality of synthetic sound was obtained. The quality of synthetic speech was evaluated not only by subjective listening but also by matching the spectra of synthetic speech with original speech.

e) Speech Processing

5.1.92 A One Parameter Control Gammachirp Filterbank For Auditory Models

Kiran G. V. T. V. Sreenivas, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

A modified one-parameter form of the Patterson, Irino gammachirp filters, as a front-end for auditory models, is presented here. The theta factor of the asymmetry term, which when multiplied with the gammatone filter gives the gammachirp filter, is made to vary directly with the level of the input signal. This modified gammachirp filterbank is compared with the previously introduced compressive gammachirp filters and the fit between the two is explored. The level dependent properties of the basilar membrane filtering action are also shown to be modeled well by this modified filter.

5.1.93 Analysis of Spoken Words Employing Gabor Transform

V.K. Madan, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

This paper employs Gabor transform to the analysis of speech signals. Speech signals were recorded under noisy conditions. The results of the analysis were compared with those obtained by employing short time Fourier transform (STFT). Gabor analysis gave, in general, a better spectral resolution as compared to STFT analysis of speech signals. Gabor analysis though had been applied to speech signals but is not yet as widely used as STFT analysis. It has, however, more potential for its application to speech processing than presently exploited.

5.1.94 Frequency Warped All-pole Modeling Of Vowel Spectra: Dependence On Voice And Vowel Quality

Pushkar Patwardhan and Preeti Rao, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

We address the problem of compactly representing the discrete spectral amplitudes of vowel sounds produced by a sinusoidal model. A study of frequency warped all pole model representation of spectral amplitudes has been presented. It has been generally accepted that incorporating Bark scale frequency warping in the all-pole modeling improves the perceived accuracy of the modeled sound. However our study suggests that whether such frequency warped all-pole modeling would improve the modeling accuracy depends on the nature of the vowel as well as the voice. We propose an alter-native warping function which may be used to improve the modeling accuracy more universally.

5.1.95 Acoustic Measures In The Speech Of Children With Stuttering And Normal Non Fluency - A Key To Differential Diagnosis

Prakash B., *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

Literature surveying various perceptual protocols and objective measures reveal no clear agreement to decide transient periods of dis/dys fluency as pathological stuttering or normal non fluency. Part of this confusion arises because of lack of objective dependable measures to differentiate the two conditions. In this context, the present study evaluated speech of 10 normal and 10 stuttering children speaking Kannada (a south Indian language) on refined acoustic measures viz. formant patterns, speed of transitions, F2 transition duration, and F2 transition range as possible indicators for differential diagnosis. Results reveal stuttering children exhibited longer transition duration, shorter extent and faster speed of transition and abnormal F2 transition patterns. The results are discussed from theoretical and clinical perspectives.

5.1.96 Begin-End Detection Using Vowel Onset Points

S.R.M Prasanna, JM Zachariah, B Yegnanarayana, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

This paper proposes a method for detecting begin and end points of a speech utterance using the Knowledge of Vowel Onset Points (VOPs). VOP is defined as the instant at which the onset of vowel takes place. An algorithm for VOP detection in continuous speech is discussed. VOP helps in overcoming the difficulties present in coming up with multiple thresholds followed in most of the existing begin-end detection algorithms. The VOP of the first vowel is used as an anchor point for further analysis to detect the begin of the speech utterance. Similarly, the VOP of the last vowel is used as an anchor point for detecting the end point. The performance of the proposed begin-end detection algorithm is compared with the existing energy-based approach by conducting text-dependent speaker verification system using the knowledge of VOP for begin-end detection shows a significant improvement in the performance.

5.1.97 Voicing Perception In Patients With Cerebellar Pathologies

S. R. Savithri & H. Rohini, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

The present study investigated the role of cerebellum in voicing perception in patients with cerebellar pathologies. Ten patients with cerebellar dysarthrias in the age range of 24-69 years participated in the study. The perception of voicing for VOT and closure duration continuum was investigated. VOT continuum consisted of 223 pairs of synthetic stimuli and closure duration continuum consisted of 150 pairs of synthetic stimuli. The patients were to respond to the stimuli on a binary forced-choice with identifying the pairs in the stimulus as “same” or “different”. The results indicated that the shift in the percept of voicing was delayed in patients with cerebellar pathologies for VOT and no shift in percept was noticed for closure duration. The results indicate that the perception of linguistically relevant intervals is impaired in individuals with cerebellar pathologies. The findings support the hypothesis that the cerebellum represents an “internal clock”, a pre-requisite for temporal computation in the perceptual domain.

Sirasta: Ivantargatam sarvendriyaparam manah tatrastham tadvisaya:nindriya:na:m rasa:dikan-sami:pasthan vija:na:ti tri:n bha:vana:nsca niyaccati Lying between the hard palate and the skull, the indispensable structure, the brain, describes the senses like taste and controls the three radical emotions. *Bhe:la Samhita*

5.1.98 M-ary Predictive Coding: A Nonlinear Model For Speech

M.Vandana, *Workshop on Spoken Language Processing, TIFR, Mumbai, January 9-11, 2003.*

Abstract

Speech Coding is pivotal in the ability of networks to support multimedia services. The technique currently used for speech coding is Linear Prediction. It models the throat as an all-pole filter i.e. using a linear difference equation. However, the physical nature of the throat is itself a clue to its nonlinear nature. Developing a nonlinear model is difficult as in the solution of nonlinear equations and the verification of nonlinear schemes.

In this paper, two nonlinear models– Quadratic Predictive Coding and M-ary Predictive Coding- have been proposed. The equations for the models were developed and the coefficients solved for. MATLAB was used for implementation and testing. This paper gives the equations and preliminary results. The proposed enhancements to MPC are also discussed in brief.

f) Typesetting

5.1.99 Indica, an Indic preprocessor for TEX

Yannis Haralambous, *1994-03-23,*
Web:http://omega.enstb.org/yannis/pdf/indica

Abstract

In this paper a two-fold project is described: the first part is a generalized preprocessor for Indic scripts with several kinds of input (LATEX commands, 7-bit ASCII, CSX, ISO 10646/ UNICODE) and TEX output. This utility is written in standard Flex (the GNU version of Lex), and hence can be painlessly compiled on any platform. The same input methods are used for all Indic languages, so that the user does not need to memorize different conventions and commands for each one of them. Moreover, the switch from one language to another can be done by use of user-defineable preprocessor directives. The second part is a complete TEX typesetting system for Sinhalese. The design of the fonts is described, and METAFONT -related features, such as metaness and optical correction, are discussed.At the end of the paper, the reader can find tables showing the different input methods for the four Indic scripts currently implemented in Indica: Devanagari, Tamil, Malayalam, Sinhalese. The author hopes to complete the implementation of more Indic languages into Indica soon; the results will appear in a forthcoming paper.

5.1.100 Low-level Devanagari Support for Omega- Adapting devnag

Yannis Haralambous and John Plaice, *Proceedings of the August 19, 2002,*
Web:http://omega.enstb.org/yannis

Abstract

This paper presents tools (OTPs and macros) for typesetting languages using the Devanagari script (Hindi, Sanskrit, Marathi). These tools are based on the Omega typesetting system and are using fonts from devnag, a package developed by Frans Velthuis in 1991. We are describing these new OTPs in detail, to provide the reader with insight into Omega techniques and allow him/her to further adapt these tools to his/her own environment (input method, font), and even to other Indic languages.

5.1.101 TeluguTeX

Lakshmi V. S. Mukkavilli, *1991,*
Web:http://www.tex.ac.uk

Abstract

There are seven sections in this article. The first section explains the Telugu script.For someone who can speak Telugu this section should help learn Telugu script. On most of the computers there is no facility for inputting Telugu text. We propose a romanization scheme for inputting Telugu. This facilitates text entry in Telugu. This is the subject of the second section. We provide examples of Telugu text typeset using our system in the third section. In the fourth section we explain the problems involved in typesetting telugu text. The fifth section deals with implementation and the sixth section contains samples of telugu text typeset using various sizes/styles. Last section summarizes our work.

5.1.102 AroSgaon (More SGAON) 2.1

Muhammad Masroor Ali, *Sept., 1996*
Web:http://www.ibiblio.org/pub/packages/tex/languages/bengali/arosgn/arosgn.ps

Abstract

The bangla font sgaon was developed by Anisur Rahman and it has found application in various packages including itrans(Avinash Chopde). Perhaps it was the first and only bangla font in public domain. The name arosgn roughly translates to more sonargaon.One need to use the itrans package for transliteration of Roman characters.Itrans is available from ftp://chandra.cis.brown.edu/pub/itrans-4.0 plus a number of FTP sites. The fontset was developed in haste solely out of personal needs. In no way it is a polished product. Sometimes extra spaces may appear between character between a sgaon character and a arosgn character (the original sgaon font is not free from this defect). You may notice that with respect to appearance, matching between the 300 DPI font supplied with the itrans package and 300 DPI arosgn fonts is the best. So while using xdvi, dvips or dvi2ps resolution selection options may become useful. Fonts at other DPI values can also be fine tuned for proper matching.

5.1.103 Fonts- An Overview of Indic Fonts for TEX

Anshuman Pandey, *TUGboat, Volume 19 (1998),No. 2,Web:http://www.tug.org*

Abstract

Many scholars prefer TEX over the "word processors" because of the case TEX provides them in typesetting non-Roman scripts and the ability TEX has in producing well-structured documents. With the development and availability of Indic language and font packages the flexibility of the LATEX2 system, there is now even more reason for indologists to implement TEX in their work. There are roughly thirteen major Indic scripts which are used throughout South Asia and all of these major scripts can be typeset with TEX.TEX takes the user beyond such diffculties by facilitating the implementation of multiple scripts without the hassle of worrying about various fonts and their encodings, manual font switching and other such hindrances to productivity caused by common word processors.TEX systems has support for following Indic fonts:Devnag, Devnac, Devnag Pen, Sanskrit, ItxBengali, ItxGujarati, Punjabi, Gurmukhi, Washington Tamil, Telugu, Malayalam, Sinhala, Konark, Cuttack, AAI Kannada, GTibetian, Naskh, Washington Brahmi.Fonts for perso-Arabic, Oriya, Sinhalese and Tibetan are also supported.

5.1.104 Devanagari for TEX version 2.0 Users Manual

Anshuman Pandey, *23 January 2000,University of Washington, Web:http://www-texdev.mpce.mq.edu.au/l2h/languages/indic/devanagari/manual*

Abstract

The Devanagari for TEX (devnag) package enables the typesetting of the Devanagari script in TEX and LATEX .The devnag package was originally developed and released in May 1991 by Frans Velthuis. Since its release, the package has been upgraded for compliance with the LATEX standard.This document accompanies version 2.0 of the devnag package. This latest release includes the newest version of the preprocessor, style files, and fonts. Version 2.0 of devnag is the first major revision of the preprocessor since its original release. It is based on version 1.6 (there are no releases 1.7, 1.8 or 1.9), and, in addition to Marc Csernel's LATEX extensions, incorporates numerous modifications made by John Smith. The style files have been revised by Dominik Wujastyk and Francois Patte, and revisions of the font made by Anshuman Pandey. This package may be found at various Comprehensive TEX Archive Networks (CTAN) such as ftp://ftp .tex .ac .uk/ in the directory tex-archive/language/ devanagari/ distrib/. Velthuis developed the devnag package with an objective of keeping the format of the input source text as close as possible to the accepted scholarly standards for transliteration. In doing so he instituted a widely used transliteration scheme known as the 'Velthuis scheme'. The 'Velthuis scheme' has been adapted by various other Indic language TEX packages and serves as the basis of other Indic transliteration schemes.The following individuals currently maintain the devnag package.

5.1.105 The obitet-package -Writing Tibetan text for Omega

Norbert Preining, *November 24, 1997, Web:http://www.logic.at/people/preining/tex/unidoc.ps*

Abstract

This package should ease the pain when writing tibetan text for the typesetting system omega, a Unicode-capable descendent of' TEX. This package makes heavy use of Omega translation process, known as omegaTP. Included in the package there are the necessary omegaTP-files to type tibetan text in a certain transcription and after processing the file with omega getting the tibetan glyphs.Together with this package comes a font tibetan.mf for METAFONT, which has to be used.This package is capable of rendering all the basic tibetan compound glyphs (asrka) in their usual form. Complex Hindi consonant stacks are made up from the basic characters by stacking them one above the other.