

FORM – 2

THE PATENTS ACT, 1970

(39 of 1970)

&

THE PATENTS RULES, 2006

PROVISIONAL SPECIFICATION

(See section 10 and rule 13)

**SYSTEM FOR RAPID PROTOTYPING OF SPEECH
RECOGNITION APPLICATIONS IN DIFFERENT LANGUAGES**

TATA CONSULTANCY SERVICES LTD.

The following specification particularly describes the nature of the invention.

FIELD OF THE INVENTION

The present invention relates to the field of speech recognition and translation.

BACKGROUND OF THE INVENTION AND PRIOR ART

Today, interactive technologies play a key role for improving customer service. The interactive technologies like IVR (Interactive Voice Response) accept verbal user input and/or request and provide pre-recorded or dynamically generated output in response to the user's request.

Typically, the IVR applications use speech recognition systems to recognize and convert words or a sequence of words spoken by a person to machine readable form. These systems are built for a source language, for instance English, due to wider acceptability of the language and the availability of information and resources in the source language. However, with increasing acceptability of speech based solutions in various countries, where the native language is different from the source language, there is a need to convert an existing speech based application working in the source language to the target language.

Typically, speech recognition based applications require:

- (i) a speech recognition (SR) engine with acoustic models for acoustic recognition;
- (ii) a pronunciation lexicon of the words which have to be recognized;
and
- (iii) a language model.

These three components work in tandem to convert the speech to text. Converting a speech recognition based solution from a source language to a target language needs these three components to be ‘ported’ to the target language.

Although, the acoustic models are tuned for a particular language, the source acoustic models can be used to recognize speech in another language with decent accuracy if the other two components, namely, lexicon and the language grammar are addressed adequately in the target language.

Essentially, converting a speech recognition application from one language to another will need creation of a new pronunciation lexicon for the target language, which contains all the words to be recognized by the speech application and its phonemic pronunciation and language grammar of the target language.

These modifications for porting a speech recognition application in source language into a target language requires efforts equivalent to building an entire new application.

Therefore, there is a need for a system which will enable an existing application to be quickly ported and/or modified to work in multiple target languages by reusing the speech recognition engine of the existing application.

OBJECTS OF THE INVENTION

It is an object of the present invention to provide a system for enabling an existing application to be quickly ported to work in another target language.

It is another object of the present invention to provide a system for accurate source to target language transliterations and translations of a speech solution.

It is yet another object of the present invention to provide a system which will automatically generate source language phonemic pronunciations of target language words.

BRIEF DESCRIPTION OF ACCOMPANYING DRAWINGS

The invention will now be described in relation to the accompanying drawings, in which:

Figure 1 illustrates the overview of the proposed system and its interface with an existing speech recognition application; and

Figure 2 illustrates the process of generating the target language lexicons.

DESCRIPTION OF THE INVENTION

Referring to the drawings, there is disclosed a system for enabling a speech recognition system in a target language by using a speech recognition system of the source language, the accompanying drawings do not limit the scope and ambit of the invention and are provided purely by way of example and illustration.

The present invention envisages a speech recognition system in a target language. Particularly, the system envisaged by the present invention enables

porting of any existing speech recognition application in a source language to a target language, thus minimizing the development process and reusing existing speech recognition applications to recognize and translate multiple languages.

The preferred embodiment of the present invention describes the conversion of the English language speech to an Indian language (Hindi) speech recognition solution.

Referring to the drawings, **Figure 1** shows the overview of an existing SR application and its interface with the system envisaged by the present invention.

Existing speech recognition applications are built of one or more call flow units, represented by block **10** of Figure 1. Each call flow unit **10** comprises of modules for performing the below functions:

- (i) prompting the users for speaking their requests, represented by block **12** of Figure 1;
- (ii) recognizing the user's speech request, represented by block **14** of Figure 1;
- (iii) processing the recognized text to answer the user, represented by block **16** of Figure 1; and
- (iv) providing response to the user, represented by block **18** of Figure 1.

There are two types of data that needs processing: (a) the speech (acoustic) data and (b) the textual data. The speech data is used at the point of

interaction with the user, while the textual data is processed internally for processing information extracted from the speech data.

The invention achieves the objective of converting an existing speech application into a target language by following the steps as given by:

1. keeping the textual data, and its processing by the existing application, unchanged;
2. using the same textual data representation (in source/English language) of the speech data, irrespective of the target language; and
3. modifying the source SR resources i.e. phoneme lexicon and grammar to process the target language.

In accordance with the present invention, the call flow units **10** along with the application data represented by block **22** of Figure 1 remains unchanged in the target language. The modification performed by the present invention is represented by block **20** of Figure 1. The present invention modifies the phoneme lexicon and the grammar for porting any existing SR (Speech Recognition) applications to a target language efficiently.

In accordance with the present invention, the Lexicon and Grammar modifications, performed by the present invention are achieved by the Lexicon Modification Engine (LME), represented by block **24** of Figure 1 and Grammar Modification Engine, represented by block **26** of Figure 1.

Figure 2 shows the detailed procedure for generating the source language phoneme pronunciation lexicon for the target language words in the application.

Lexicon Modification Engine:

Lexicon Modification is the main and often the only required modification to the existing SR applications. In accordance with the present invention, the Lexicon Modification Engine (LME) **24** involves automatic creation of the pronunciation lexicon represented by block **28** of Figure 1, of the words to be recognized, which are present in the source language, into the target language. LME **24** consists of a source to target word dictionary, which is a database of words for translating each word in the source lexicon to the target grapheme. LME **24** receives data/words from Application Data **22** and gives the pronunciations to Phoneme lexicon **28**.

The LME **24** takes each word in the source lexicon represented by step **102** of Figure 2, and determines its translation from the source language to target language using the word dictionary represented by block **100** of Figure 2, in the target grapheme, for example gold is converted to सोना represented by step **108** of Figure 2, then a transliteration is done to convert source language सोना to sona represented by step **110** of Figure 2. The pronunciation is determined from sona as “s ow n aa” using a grapheme to phoneme in the source language represented by step **112** of Figure 2 and as seen in Table 1. This process speeds up the process of lexicon creation in the target language though the grapheme to phoneme conversion in the source language for a word in the target language. The ‘Phoneme Lexicon’ **28** provides the pronunciations to **14** of the call flow unit **10**.

	English	Hindi
Grammar phrase	< Gold >	< Gold >
Lexicon entry	/g/ow/l/d/	/s/ow/n/aa/

Application asks for input	- (grammar unchanged)	- (grammar unchanged)
User speaks	<i>/gold/</i>	<i>/sonaa/</i>
SR output and Process input	‘Gold’	‘Gold’

Table 1

If the source language word has a translation in source to target language dictionary **100**, then the target language word is fetched, as represented by path referenced by numeral **106** in Figure 2 otherwise path represented by reference numeral **104** in Figure 2 is taken, where the source language word is transliterated to target language (assuming it is spoken in the same way as in the target language e.g. proper nouns).

In accordance with another aspect of the present invention, source to target to source transliteration is done because direct source to phoneme conversion in the source language may not give correct pronunciation phoneme sequence of the target language word.

Grammar Modification Engine:

Grammar modification is generally not required for an existing menu driven based speech application because the application expects only a word or a small sequence of words as the input from the user. These words are addressed by the source-target word dictionary to create the pronunciation lexicon **28**.

Grammar modification (source-to-target) is required in cases where the speech application system is expected to handle free speech queries. The

grammar creation for the target language is achieved by source to target language translation followed by a transliteration to source language.

The Grammar Modification Engine **26** performs these translations and transliterations and gives the output as grammar represented by block **30**.

If recorded prompts are used in the existing application, then a similar database of prompts in the target language is created and the application points to this database for prompts and responses to the user.

The technical advancements of the present invention include:

- providing a system for multilingual speech recognition;
- providing a system for enabling an existing application to be quickly ported to work in another language;
- providing a system for accurate source to target language transliterations and translations;
- providing a system which will generate source language phonemic pronunciations of target language words;
- providing a system which minimizes the efforts equivalent to designing a new application in the target language when wanting to ‘port’ the existing application in a source language in target language; and
- providing a system which reuses the original application and business logic.

While considerable emphasis has been placed herein on the particular features of this invention, it will be appreciated that various modifications can be made, and that many changes can be made in the preferred

embodiments without departing from the principles of the invention. These and other modifications in the nature of the invention or the preferred embodiments will be apparent to those skilled in the art from the disclosure herein, whereby it is to be distinctly understood that the foregoing descriptive matter is to be interpreted merely as illustrative of the invention and not as a limitation.

Dated this day of May, 2009

MOHAN DEWAN

Of R. K. DEWAN & CO.

APPLICANT'S PATENT ATTORNEY