

ASR Phonetic Learning Technology

The most critical aspect in speech applications is the need to correctly predict user formulations and to effectively deal with non-native speakers or regional accents. In a directory assistance application, for instance, user formulations for business listings, may differ a lot with respect to the system knowledge derived from white and yellow pages. In a very complex application, even if a careful study of user behavior has been performed, it is impossible to predict exactly how users will formulate their requests. The knowledge provided to the system by a priori analysis is very useful in order to release the first version of a speech application, but is nevertheless not enough. The fundamental source of information is the data the system collects from the field. The processing of very large amounts of collected data, however, can prove too expensive if performed by human operators. This is why Loquendo has developed a new technology, which automatically analyzes application data to detect the most significant weaknesses.

Loquendo ASR Phonetic Learning technology is available to speech application developers to improve application performance. It addresses two main issues, which aim to improve the effectiveness of the grammar recognition objects:

- Automatic discovery of pronunciation variants for the vocabulary words
- Clustering of frequent unforeseen linguistic formulations

Loquendo Phonetic Learning Algorithms

Today, speech applications are mainly based on recognition grammars. A recognition grammar is a formalism that allows application designers to specify what a user can say. The recognized sequence of words must be foreseen within the grammar. However, a user might utter a sentence that is not completely covered by the grammar. Moreover, the speaker's pronunciation may not be completely covered by the system's phonetic knowledge, mainly in the case of foreign words. When these phenomena become frequent, they can become problematic from the application point of view. How can one detect them? The stored application log data contains all the information related to a single recognition interaction, including the recognized words, their confidence values, the audio signal and the phonetic transcriptions decoded through a phone network. It is quite likely that a poor confidence score triggers several mismatched conditions. Different repetitions of the same utterances should provide similar phonetic transcriptions. Loquendo phonetic learning technology is capable of finding clusters characterized by phonetically similar utterances. A representative phonetic transcription for the most populated clusters is provided.

An artificial example of the phonetic learning process is shown in Figure 1. In this case, a date grammar has been designed without the inclusion of the month of "April", so a systematic error occurs when the word "April" is uttered. However, the system may detect recognized speech portions characterized by low confidence values. The phonetic transcriptions associated to these portions can contribute to the building of a new cluster if other similar transcriptions are frequently detected in the log data.

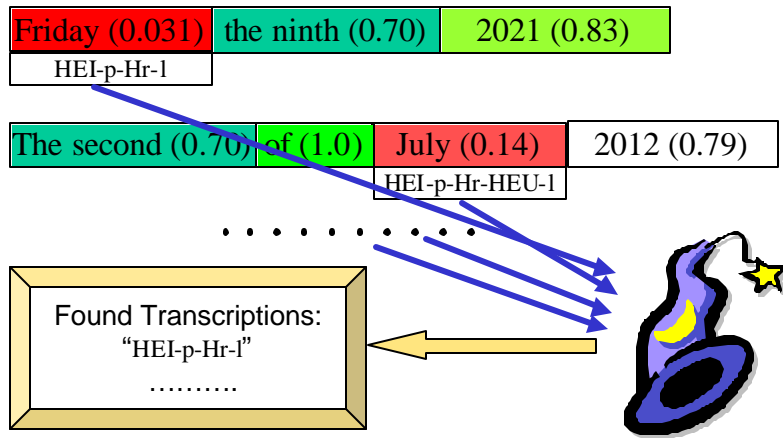


Figure 1 Phonetic learning processing

The same approach is used when searching for pronunciation variants of vocabulary words. In this case, the cluster is made up of automatically derived phonetic transcriptions related to a given recognized word. Only repetitions with medium / high confidence score values are taken into account in order to obtain high reliability on the recognition results. Additional phonetic transcriptions can be found and they can be added to the canonical ones.

Loquendo phonetic learning architecture

The Loquendo ASR phonetic learning architecture is outlined in Figure 2.

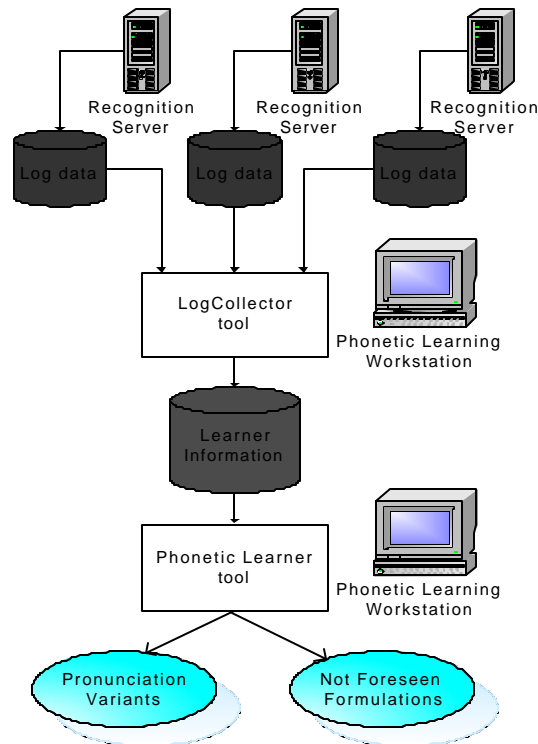


Figure 2 The phonetic learning architecture

In order to use recognition data collected from the field, it is necessary to save it. Loquendo ASR log mode allows the automatic saving of recognition information. It is possible to enable log features with or without dumping audio data, and optionally to activate the production of phonetic learning related specific information (phonetic net decoding). This data is processed for phonetic learning purposes only, and employ quite relevant CPU and memory hardware resources. For this reason, we suggest that the production of this information be enabled only when phonetic learning is required.

Typically, a voice application may need more than a single recognition server to be available to the users. The Loquendo ASR log will be produced on a recognition server basis, but it is convenient to merge all the collected information, because the phonetic learning algorithms require a large number of data to increase the statistical significance of the produced results. The Loquendo LogCollector tool gets log information data from different recognition servers, producing a unique local database that is usable for phonetic learning purposes. The LogCollector tool is displayed in Figure 3.

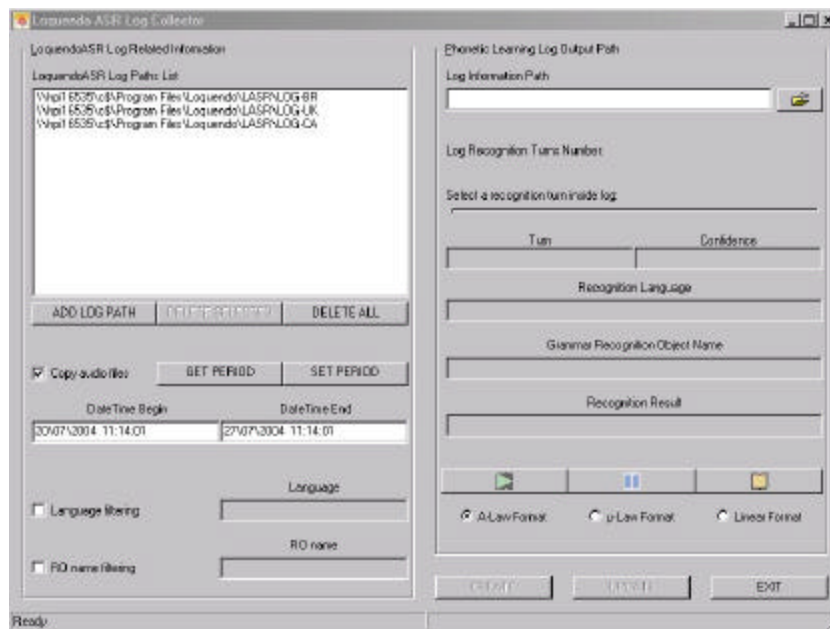


Figure 3 The LogCollector tool

The last step for performing phonetic learning analysis, is the use of data saved by the LogCollector software inside the Learner tool. The Learner tool will produce a hypothesis related to pronunciation variants or unforeseen formulations. The application developer should check the produced hypothesis, and accept or refuse them. If audio dumping has been enabled, the developer may also use the audio information to check the phonetic learning results on the basis of an immediate match between phonetic transcriptions and audio. The Learner tool is displayed in Figure 4.

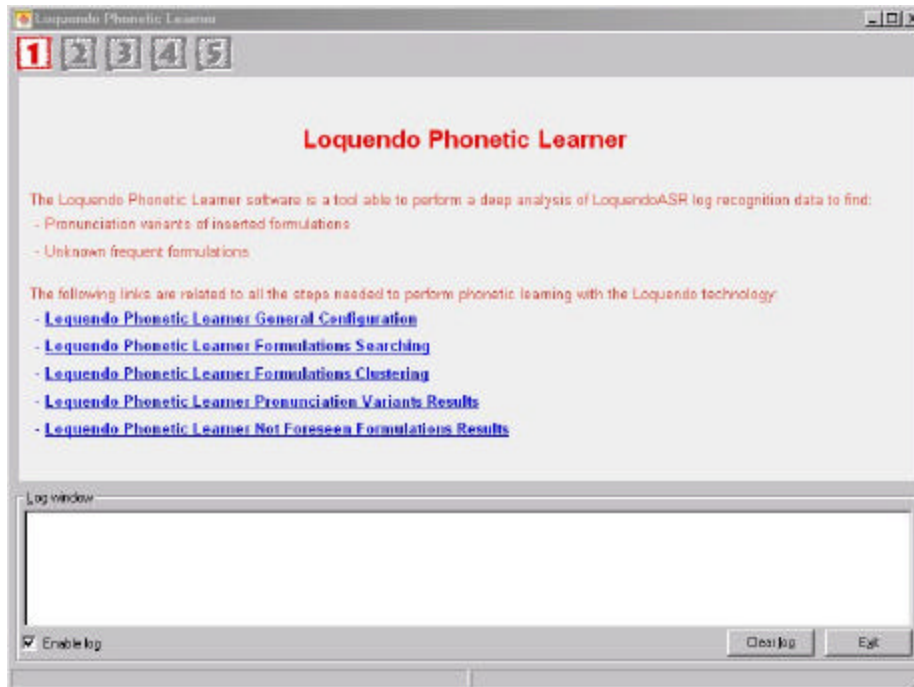


Figure 4 The Learner tool

Loquendo phonetic learning technology is currently available for all Loquendo ASR languages, with the exception of Greek and mixed Castilian - Catalan.

The Loquendo phonetic learning experience

Loquendo has applied phonetic learning technology to a fully-automated directory service developed by Loquendo for Telecom Italia. This service has been live since 2000. Given a Name and Address, it returns a Phone Number from a database of 25 million Italian subscribers. The automatic system can fully automate both business and residential requests. All calls are routed to the automatic system; the calls that cannot be automated are routed to human operators, together with transcribed data.

Directory Assistance for business listings is a challenging task: one of its main problems is that customers formulate their requests for the same listing with great variability. Since it is difficult to reliably predict user formulations *a priori*, Loquendo has studied a procedure for detecting, user formulations that were not foreseen by the designers from the field data itself. These formulations can be added, as variants, to the denominations already included in the system in order to reduce failures. In particular, it is fundamental to detect new formulations for frequently requested listings, for example “nicknames” of hospitals or other public services, or user requests for the phone number of a popular TV talk show, that of course do not appear in the directory listings. Loquendo approach is based on partitioning the field data into phonetically similar clusters, from which new user formulations can be derived.

The results of the experiments on a very large amount of calls that the system was unable to serve automatically, are very positive indeed. The technology allowed considerable improvement in terms of linguistic coverage, reducing the mismatch among the most frequently asked business listings and the system’s knowledge. The system is now operative, and allows periodical updates to the formulation variants system for each town.