WHITE PAPER

NSC
Natural Speech Communication Ltd.

July, 2005

Key-Word Spotting- The Base Technology for Speech Analytics

By Guy Alon
NSC - Natural Speech Communication Ltd.

33 Lazarov St. Rishon Lezion Israel
Tel. +972-3-9519779
e-mail: info@nscspeech.com

1. Introduction

With the development and worldwide adoption of Automatic Speech Recognition (ASR) technology, additional ASR markets and applications are emerging with specific needs from the ASR engine.

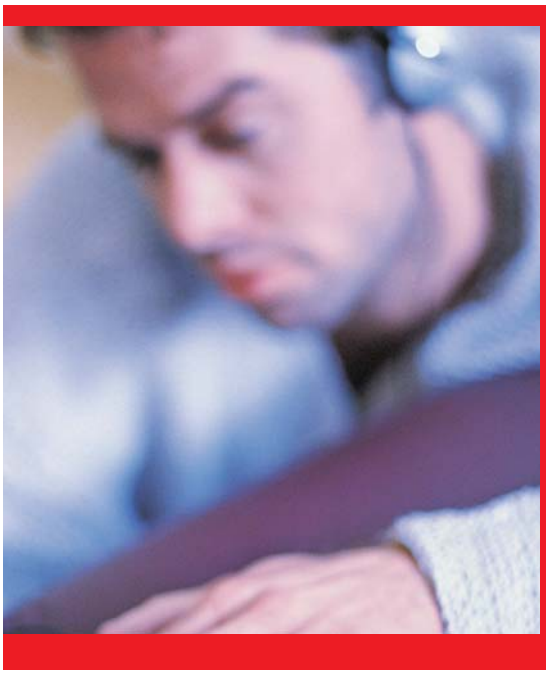
A new emerging application in which ASR engines can be used is automatic call analysis. Such an application usually includes: speech recognition, speech analysis and data mining as well as other base technologies and tools. In such an application, the speech recognition engine is used for key-word spotting and provides the input for advanced surveillance and analytical application.

Key-Word Spotting (KWS) - the ability to use the ASR engine in order to analyze natural day-to-day conversation between two or more people and indicate whether specific key-words were mentioned during the call.

Such a capability can have great importance for mining valuable information from a mass number of telephone calls. The ability to dig information out of the **call contents** automatically, and not only data from the Call Detail Records (CDR) represents an innovative approach in analyzing caller's behavior, intention and state of mind and has great value. Although the concept of KWS is not new, the ability to deliver such a functionality for large-scale systems with acceptable performance and accuracy has only become available in recent years.

In order to overcome the challenges of delivering a robust KWS product, several approaches have been adopted by the vendors, each with its own unique characteristics.

This paper will review the KWS challenge, the main requirements of the various target market from the KWS engine and the different approaches of delivering a KWS engine.



2. The KWS Challenge

The technology of searching key-words in audio streams using an ASR engine has particular characteristics that make it different from telephony voice-driven services and a much more difficult task.

In typical voice-driven services, the content of the speech is context-dependent (e.g. answer to a specific question or a sentence within a known domain) and the speech duration is normally a sentence or phrase of a few words or less. Unlike voice-driven services, the KWS problem deals with continuous spontaneous conversation between two or more speakers that can include an unlimited range of vocabulary. In addition, the speech can be from a noisy environment, various input types and include several speakers.

KWS has the ability to deliver functionality for large-scale systems with acceptable performance and accuracy and has only become available in recent years.

Another unique challenge is that KWS applications typically need a larger number of ASR channels than in human-machine interaction since the KWS channels are allocated to the entire duration of phone conversations. In addition, the number of calls to be analyzed using a KWS product is huge compared to the number of those in speech-driven calls.

3. The KWS Target Market Segments

Different potential market segments for key-word spotting each have a different set of expectations and needs from the ASR solution. This section will review the main market segments for KWS and their unique requirements from the engine.

When looking at the potential market segments for KWS, the two most natural segments are **Homeland Security (HLS)** and **Call Center (CC)** markets, although other market segments analyzing large numbers of speech recordings can be targeted as well.

The Homeland Security Market

It is no secret that since the tragedy of September 11 and other acts of terror in other geographical regions, the investment in homeland security had been dramatically increased. Research conducted by various industry analysts allocates figures ranging from 22 billion dollars a year up to 72 billion dollars a year. Part of this budget includes allocations for advanced surveillance and monitoring systems and technologies.

The homeland security market seems to be the one that is most willing and open to adopt key-word spotting, mainly due to governmental pressure, the increasing threat of worldwide terrorism and the inability to process such huge amounts of data by human resources.

The main motivation would be to search for specific key-words in a huge number of phone conversations automatically in order to decide which calls should be monitored by human listeners.

A very unique need for this market is the ability to track key-words from any type of speech stream, not only in telephony. The need also exists to apply KWS capabilities on cellular, radio and video audio streams.

The Call Center Market

The second potential market segment for KWS is monitoring and analyzing calls in the call center. Some motivations to use KWS are:

Agent supervision and quality assurance: Call center supervisors are often required to listen to on-going calls as part of the quality assurance process in the call center. The goal is to assure that the agents are properly trained with the company's messages and portfolio, and assure that agents know how to manage specific situations with customers.

KWS can be used here in order to allow the supervisor to join "problematic" calls in real time. In some call centers, the supervisor is equipped with the means to instruct the agent via chat of how to deal with the situation, while in others the supervisor can barge into the call. In any case, the ability to handle these situations as they appear is important in reducing customer churn and increasing customer satisfaction.

Business intelligence: Using key-word spotting, it is possible to analyze agents' performance, as well as categorize spotted calls according to agent name, type of problem or key word, discussed products or services, etc. This wealth of information is highly valuable to the business since it offers means for analyzing customers' responses to the service and monitoring the overall company performance.

KWS information can be used in analytical CRM applications to analyze agent-customer calls and gain important business insight for customer scoring, campaign effectiveness surveys, customer satisfaction index and other valuable business information.

Additional Markets

Additional markets that might consider adopting key-word spotting are the Telco, finance, insurance, broadcasting and healthcare markets. In fact, any market that requires analysis of large amounts of speech can be considered a potential market.

4. KWS and Speech Analytics

"Speech Analytics" is a fast growing niche in speech applications. The concept of **speech analytics** is to use various core technologies and develop targeted applications in order to mine valuable information from monitored calls.

In order to do so, the Speech Analytics vendors need to develop advanced applications that use the KWS engine's capabilities. As described above, the target markets, at this point in time, are primarily the security and the call center markets. The Analysis requirements for the homeland security market mainly require the following three capabilities: 1) the ability to support specific languages (not only at the recognition level, but also at the analytical level); 2) the ability to provide fast processing, or even real-time, analysis; 3) the ability to flag suspicious calls in a large amount of speech data.

Other requirements at the analysis level that are common to both markets are:

- To find hidden trends and patterns in the contents of the calls using data mining techniques.
- Intelligent categorization of incoming calls according to any pre-defined criteria (example: specific competitors or products, service cancellation requests, balance queries, etc.).
- High-level management reporting.
- High-level monitoring of the call center performance via dashboards and report menus.
- To find correlation between different types of calls and events.
- The ability to run ad-hoc searches over a recorded call database.

As mentioned before, every speech analytics vendor relies on a KWS engine which gets the voice inputs, and provides different outputs as discussed below.

A good KWS Engine is the key to success to every speech analytics project.

5. Main Requirements from the KWS Engine

KWS Engine Performance Measures

KWS engine performance can be defined by several measures:

- **Correct spotting** - the percentage of correct key-words spotted.
- **False reject** - the percentage of target key-words falsely rejecte.
- **False alarm** - the number of times a word was falsely spotted in a given time frame.

The main requirements from a KWS engine are described below:

Performance: As described in the previous paragraph, KWS engine performance can be defined by several measures. **Since the decision is threshold-based, a trade-off exists between false alarms and false rejects.** The priorities of the various performance measures can be different for each market and/or customer but in general, the accuracy should have acceptable rates so that a meaningful percentage of the results is useful.

Processing time: Processing time becomes a factor due to the large amounts of data, the need to process it in a given time frame and the number of available KWS channels. In some cases (e.g. homeland security), it may be important to have an immediate indication of imminent security threats, in which case KWS may be needed to operate **in real time**.

Real time processing is an advantage since it can be deployed 24-hours a day and provides 1:1 ratio between the actual call duration and the time required for the anaylsis.

This is also beneficial when off-line processing is done, since it might require fewer KWS-channels when many hours of monitoring are required.

Robustness: The need to monitor a large number of parallel calls requires a robust and scalable solution. Therefore, a high-density solution is required, one that has the ability to perform the call monitoring in a given reasonable physical space along with computing resources for thousands of concurrent calls.

Flexibility: Flexible and easy to edit tools are required for independently editing the key-word list. The ability to fine-tune the sensitivity of the KWS engine is required in order to optimize performance in a specific working environment and according to customer needs.

Language support: The KWS engine should have the ability to support multiple languages within the same system. In the homeland security market, support of more remote languages is required.

Architectural aspects: Typically, a KWS engine would be added to an existing system, which usually includes a recording platform. An easy integration process with minimum influence on existing system architecture and resources is desirable.

Real time processing is an advantage since it can be deployed 24-hours a day and provides 1:1 ratio between the actual call duration and the time required for the anaylsis.

6. Approaches in Implementing KWS

Various approaches exist for implementing a KWS engine. There are basically three underlying processes used:

- **LVCSR based KWS** - This approach uses a two-stage process. In the first stage, the transcription of the speech into words is done using a Large Vocabulary Continuous Speech Recognition (LVCSR) engine, outputting formatted text. In the second stage, a textual search for the key-words within the text is performed. Using this approach, results from LVCSR and the text search are combined to spot the key-words.
- **Phoneme Recognition based KWS** - This approach also uses a two-stage process. In the first stage, the speech is transformed to a sequence of phonemes. In the second stage, the application searches for phonetically transcribed key-words in the phoneme sequence obtained from the first stage.
- **Word Recognition based KWS** - This approach searches for the key-words in a one stage operation. The recognition is phoneme-based and the KWS engine looks for the key-word in the speech stream based on a target sequence of phonemes representing the key-word.

7. Comparison of KWS Approaches

In this section, the three approaches are compared and the advantages and disadvantages of each approach are discussed.

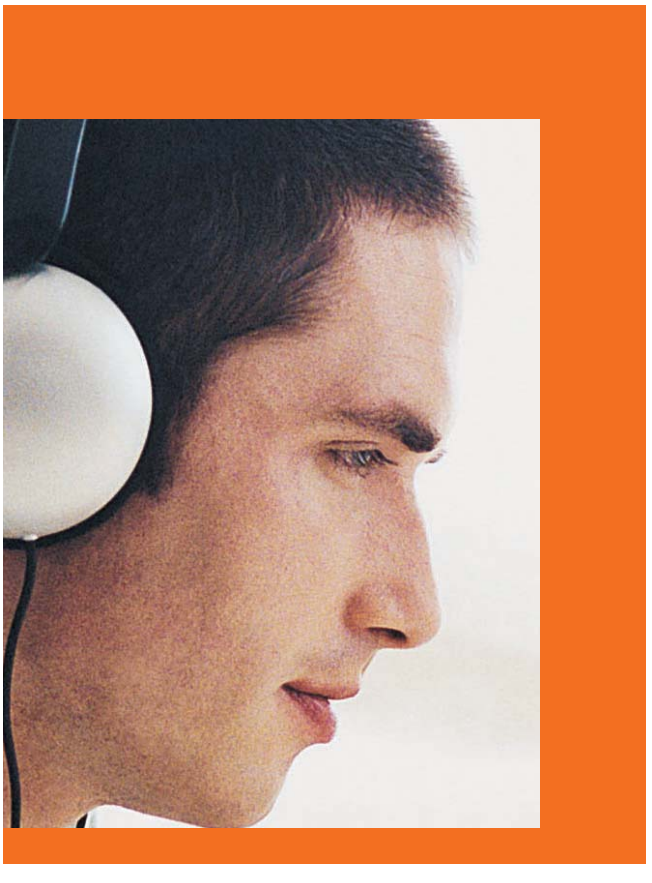
- **LVCSR based KWS -** This approach is suitable for searching over databases in which you cannot run the KWS engine. In this case, the very large speech database is transformed into text using a LVCSR engine for later faster textual search.

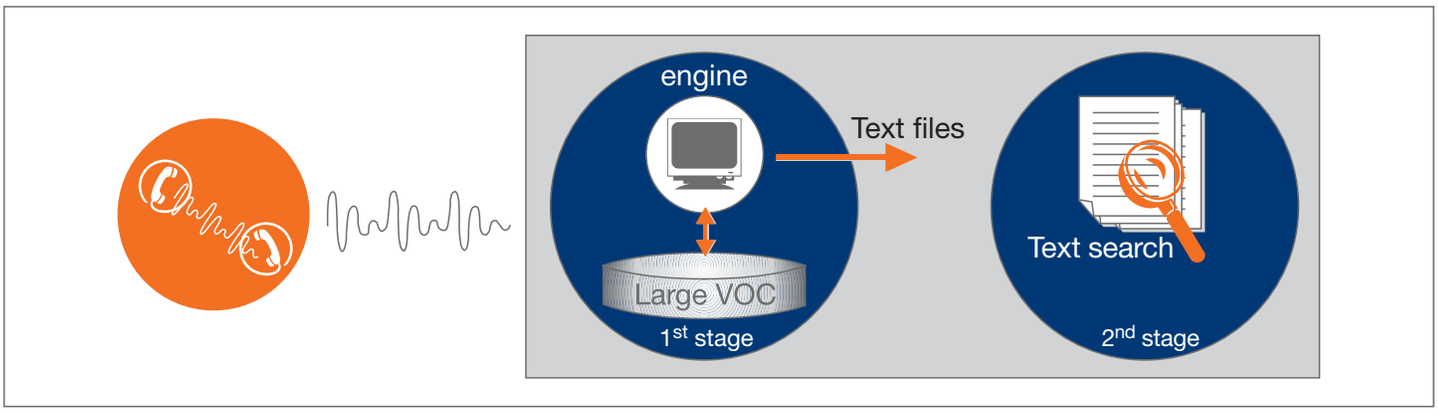
The main advantages of this approach are:

1) Easy and economical storing of call transcriptions
2) Fast re-search of the same key-words
3) Key words do not have to be pre-defined
4) Once the transcription is done and assuming it is of high quality, searching for new keywords in the existing text is straightforward and faster than a speech search.

The main disadvantages of this approach are:

1) The usage of LVCSR cannot cover the complete set of words that might represent key-words, especially foreign names and terms. The addition of new words to the vocabulary is complex
2) LVCSR is a heavy consumer of CPU and memory resources; thus, it is limited in its ability to support large-scale systems when short processing time is required.





LVCSR based KWS Diagram

- **Phoneme Recognition based KWS**

  As opposed to the LVCSR approach, this approach works at a lower linguistic level - the phoneme level.

  Phonemes are the basic speech units that, when combined, can represent any word in the language.

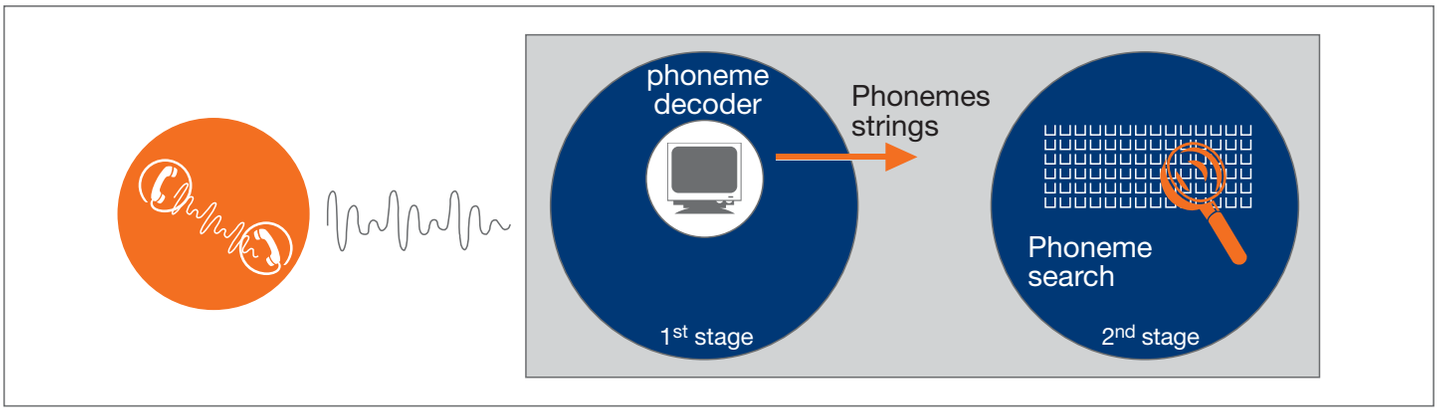
  In the first stage, the speech is also transformed into text, but this time the text is a sequence of phonemes, rather than a sequence of words. In the second stage, a search is done for a sequence of phonemes that represent the key-word.

The main advantages of this approach are:

1) Second-stage search is faster than searching directly on the speech stream.
2) List of key words does not have to be pre-defined

The main disadvantages of this approach are:

1) The first stage of phoneme recognition generates a "noisy" result, where the error rates are not minor, and can affect the KWS results performed at the second stage.
2) The phoneme-search for key-word representations typically requires more resources than most text searches used in the LVCSR approach.
3) This approach contains two stages, which requires massive processing power and limit it from being able to support real-time or high densities in a large scale project.



Phoneme Recognition based KWS Diagram

- **Word Recognition based KWS**

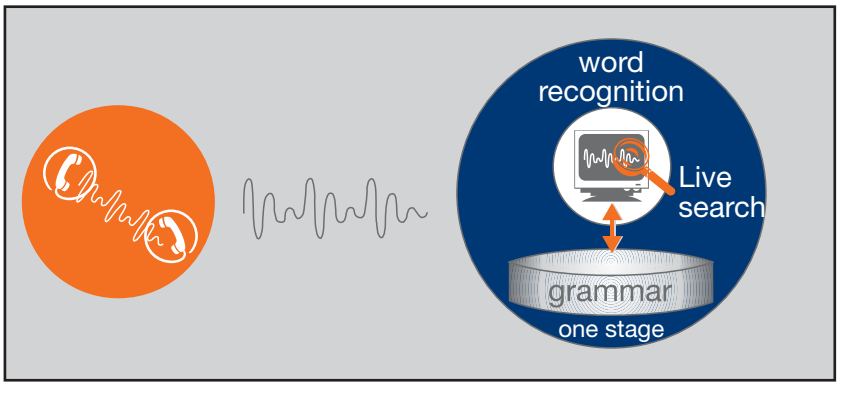
  In this approach, used by NSC, a phoneme based KWS engine is implemented. The full search of the list of key-words is done in one stage, with no preprocessing of the audio stream into words or phoneme sequences. The KWS engine receives as its input the textual list of words, represented by phonemes, and delivers at its output the recognized key-words.

The main advantages of this approach are:

1) This one-stage operation can be done in real time, but can also be done off-line if desired.
2) Since the actual specific words are defined in advance and the search is very constrained, it may lead to better KWS performance than the other approaches.
3) KWS of uncommon names and words, such as foreign names or other OOV (out of vocabulary) words is easily handled.
4) This approach is much more suitable to support the search of key-words in multiple languages in the same call.

The main disadvantages of this approach are:

1) Any re-search or change in the key word list will require a re-run of the audio stream.
2) List of key-words has to be defined in advance, but can be easily changed.



Word Recognition based KWS Diagram

8. Implementing KWS on PSP based PCI Blades

Until now, the discussion has focused on the main approaches to implementing a KWS engine. If we take a look at the platforms that the engines run on, we can distinguish between standard industry platforms on which the software-based engines are running, and specialized platforms.

Designing special PCI blades on which the KWS engine is implemented includes having many DSP chipsets over standard PCI blades with all the peripheral resources located on the PCI blade, generating a platform with a standard interface. **This represents a very unique offering.**

Using such a PCI blade platform leads to:

- A standard blade platform with PCI interface that can be plugged into any industry standard platforms.
- The KWS engine operation is done on its dedicated platform, with no need for additional CPU and memory resources.
- Such a dedicated solution is built with very high-density capabilities required for large deployments.
- Since the solution only needs free PCI slots in the system and is a high-density solution, it has a smaller impact on the existing system architecture - no servers are needed to run the KWS engine.
- The hardware platform is robust and scalable.

Additional advantages of the “Word Recognition KWS” approach using such a platform are:

- The platform can communicate directly with the telephony or recording boards working in real time on the incoming speech.
- Hundreds of concurrent channels can be monitored simultaneously, with flexible handling of language and vocabulary needs.

Using a PCI blade based engine to perform the KWS action enables overcoming the main obstacles of KWS as described in this paper.

9. Summary

The KWS market is an emerging market with unique characteristics and requirements. The emerging needs, especially in the homeland security and call center markets requiring speech analytics, can be met by using KWS technology. NSC believes that in the near future, KWS will be widely adopted by target markets and will reach full maturity.

About NSC

NSC is a leading vendor of PCI blades based ASR engines for the telephony market. The engine can be operated as a key-word spotting engine, a phoneme recognizer or speech recognition engine for speech-driven services. NSC's solution is based on PCI blades that are designed around dedicated processors (DSPs).

NSC's unique approach allows very high densities at low cost with simple system architecture. NSC's PCI blade based solution performs speech recognition without any need for CPU resources, while maintaining minimal footprints, side by side with maximum control for the users in optimizing system resources for their needs.

The company is active in the US, Europe and Middle East and involved in the implementation of ASR-based applications through business partners: IVR and VXML platforms, telephony application developers, system integrators, recording platforms developers and customers such as cellular network operators.

About the Author

Mr. Guy Alon is the Marketing Manager of NSC.

Mr. Alon has gained more than 13 years of experience in Product Management and other marketing positions. Previously to working at NSC, Mr. Alon worked for Amdocs as a Marketing Manager. Mr. Alon holds a B.Sc. in Mathematics and an MBA in Marketing.

Head Office: 33 Lazarov St., P.O. Box 5212 Rishon LeZion 75150, Israel Tel: +972-3-951-9779 Fax: +972-3-951-9671

UK Office: Electronics 2000 Ltd., High Wycombe HP12 3AJ UK Tel: +014-94-444-044 Fax: +014-94-470-499

USA Office: 30 West 21st Street, 10th floor, New York NY 10010 Tel: 1-800-238-6768 Fax: +1-212-798-1461

Germany Office: Germeringerstr. 5 Gauting 82131 Germany Tel: +49-89-891-36495 Fax: +49-89-891-36499

www.nscspeech.com