

A light green silhouette of a person's head and shoulders, facing right. The person is wearing a headset with two earpieces. The silhouette is positioned on the left side of the page, partially overlapping the text area.

WHITE PAPER

## **NetPeople<sup>®</sup>** and Enhanced IVR Systems

April 12, 2006



## Contents

Introduction .....	3
Traditional IVR Limitations .....	3
Current Speech Recognition Solutions .....	3
The NetPeople Advantage .....	3
Generated Grammars .....	3
Context Sensitive.....	4
Concept-Based Grammars.....	4
Dictation and Language Models.....	4
NetPeople Content .....	5
NetPeople Logs.....	5
Conclusion .....	5

## Introduction

IVR systems have long lived up to the “interactive” part of their name, but the “voice recognition” has until recently been weak. Speech recognition is finally becoming a viable alternative to DTMF<sup>1</sup> input for getting data from users in these applications. By using better matching algorithms, better noise cancellation and more powerful processors, companies such as Nuance® have made great strides in the recognition capabilities of IVR systems.

## Traditional IVR Limitations

One of the most frustrating aspects of IVR for users is the limited number of choices that they get at each linear step of the systems. Traditionally, IVR systems dictate a menu of options available and require the user to select the [DTMF] key associated with their desired option. This limits the complexity of processes that can be handled via IVR. The advent of usable voice recognition freed users from the necessity of pushing buttons, but did not significantly reduce the reliance on strictly linear process flows.

## Current Speech Recognition Solutions

New technologies, such as Nuance’s Say Anything™, AccuRoute™ and OpenSpeech™ Dialog<sup>2</sup>, attempt to broaden the range of options available at each stage by providing “short-cuts” to appropriate functions. Nonetheless, they still rely on structured call flows that require extensive work to develop and maintain.

## The NetPeople Advantage

By its very nature, NetPeople provides a much more human-like model of conversations, where the system can understand virtually anything that’s said within its domain at any time. Unlike conventional IVR systems, NetPeople maintains a dynamic model of the conversation as it progresses, allowing it to anticipate likely courses of conversation without constraining the user to follow them. Using NetPeople allows solution developers to concentrate on the content, and not the underlying system that attempts to interpret and act on user input.

## Generated Grammars

Along with the growing sophistication of speech recognition technology comes ever increasing complexity in the grammar syntax<sup>3,4</sup>. In order to allow grammars to enable the full breadth of options and sophistication of the latest engines, the process of creating grammars has become a long and involved endeavor requiring great expertise. Although NetPeople’s internal representation of conversations does not mirror the design used for creating grammars, the system nonetheless has all of the information necessary to create even the most complex grammars accurately. The NetPeople Publisher can

generate all the grammars necessary for implementing its sophisticated conversation model in speech recognition.

Not only does NetPeople free content developers from having to create grammars, it has two important features that are significant improvements over non-NetPeople systems: context sensitive grammars, and concept-based grammars.

## ***Context Sensitive***

Part of developing a NetPeople agent involves creating a tree of contexts, or topics, into which content is placed. At its most basic, this helps NetPeople understand pronouns and ambiguous statements. However, context-based grammars also allow NetPeople to prioritize grammar files to maximize a speech recognition engine's ability to correctly identify user statements. By understanding the current point in the overall conversation, and recent history, NetPeople can give the most relevant grammar files the highest priority proactively.

## ***Concept-Based Grammars***

Rather than just creating grammars that allow the user to speak in the way that the content creator anticipates, NetPeople uses its Natural Language Processing engine to create grammars that allow a wider variety of expressions. Not only does this cut down on the amount of work that a content creator has to do, it is more reliable than expecting the author to remember every variation of expression that could be used for a particular statement. As an example, if the user is to be able to ask the IVR for the current time, there are a number of different ways that they *should* be able to state it:

- What time is it?
- Can you tell me the time?
- Time.
- Please let me know the time.
- Do you know what time it is?

NetPeople's internal representation of all of the above examples is exactly the same, and the Grammar Generator will create a grammar that will accept any of them.

## **Dictation and Language Models**

The most flexible approach to speech recognition is called Dictation, in which the engine attempts to recognize user utterances without following any structured grammar models. Although it does indeed offer a wider vocabulary than grammar-based recognition, it suffers from a lower accuracy. One technique that has been developed to improve Dictation recognition is the use of language models<sup>5</sup>. These are derived from domain-specific corpora of text, and take the form of statistical models of word order and patterns. In a way, they can be thought of as generic grammar files that are used for all speech recognition in a given application.

### ***NetPeople Content***

Some Dictation speech recognition engines provide tools for generating language models. When such tools are available, it is possible to use special NetPeople content output as the corpora for creating very specific, well-targeted language models. By definition, the NetPeople content represents the entire domain of the application in question, and therefore is the perfect source of corpus data.

### ***NetPeople Logs***

Despite this paper's focus on NetPeople as a technology for developing enhanced IVR systems, the fact that it can be used with a wide variety of devices cannot be forgotten. In fact, using NetPeople in a text to text mode can be a very useful and time-saving method for testing IVR content. The logs generated from test cases can be used to identify holes in the content (through reports generated by the NetPeople Analytics package). However, one of the most powerful uses of the conversation logs is as an additional set of corpora for the speech recognition engine's language model tools. This can add speech patterns not anticipated by the natural language engine (which is limited to correct grammar, for instance).

## **Conclusion**

By leveraging and improving upon the current state-of-the-art in speech recognition technologies, NetPeople can bring natural conversation to a wide variety of telephony applications. As a multi-device platform, it also allows an application to be developed once and deployed across many different channels (such as phone and web), without giving up the special features required to make each experience painless for users.

---

<sup>1</sup> "Dual tone multifrequency signalling", [http://www.atis.org/tg2k/\\_dual-tone\\_multifrequency\\_signaling.html](http://www.atis.org/tg2k/_dual-tone_multifrequency_signaling.html).

<sup>2</sup> "OpenSpeech Dialog", <http://www.nuance.com/dialog/>

<sup>3</sup> "Speech Recognition Grammar Specification Version 1.0", <http://www.w3.org/TR/speech-grammar/>

<sup>4</sup> "JSpeech Grammar Format", <http://www.w3.org/TR/jsgf/>

<sup>5</sup> Ueberla, J. "Analyzing and Improving Statistical Language Models for Speech Recognition". PhD thesis, Simon Fraser University, Vancouver, Canada, 1994.