

# Intellectual Property for Speech Interfaces

VOXI OFFERS VOXIC™ – a speaker independent speech recognition and understanding interface as an Intellectual Property (IP) for embedded applications. The design is based on the same speech recognition engine used in other PC-based Voxi products. Speaker independence is achieved with use of phoneme based acoustic models, the phoneme database.

Natural language interfaces for user applications are designed with use of Voxi development tool set including the Voximizer™. The Voximizer™ has a graphical user interface whereby users can easily create any speech application.

The VoxIC™ core is intended for implementation in Field Programmable Gate Array (FPGA) or Application Specific Integrated Circuit (ASIC) devices. Care has been taken as to offer a cost effective small footprint yet powerful solution.

Possible usage includes consumer products such as remote controls and mp3 players, products for disabled people, automotive use, intelligent homes, PDAs, wearable computers and more.

## Features

- Highly parameterized core. Based on VHDL hardware description language defined by IEEE (IEEE 1076 with extension 1164).
- Supports a wide variety of target technologies including FPGAs from well-known providers Xilinx and Altera as well as ASIC implementations.
- Scripts for industry leading synthesis products from Synplicity, Synopsys and Mentor Graphics for easy implementation flow.

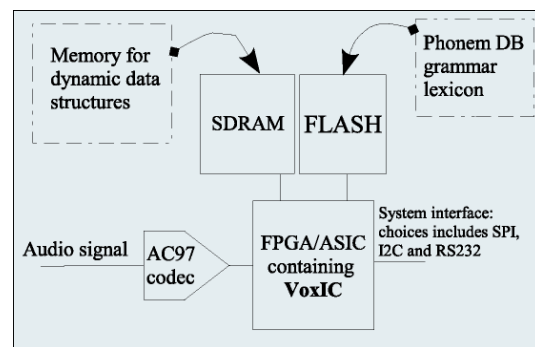
## Functional Description

The design is made up from the following conceptual components:

- Speaker independent language specific phoneme database
- Lexicon

- Grammar
- Signal processing unit
- Comparison of slices of utterance against content in phoneme database
- Hypothesis handler
- Interface to user system through user selectable peripheral bus

The core VoxIC™ dynamically creates hypotheses based on the lexicon and grammar and the evaluation of the current utterance. Below is a conceptual picture of how the core is implemented in a system.



The phoneme database, lexicon and the grammar definitions are stored in external FLASH memory. To achieve real-time evaluation the content is transferred to fast SDRAM upon start-up.

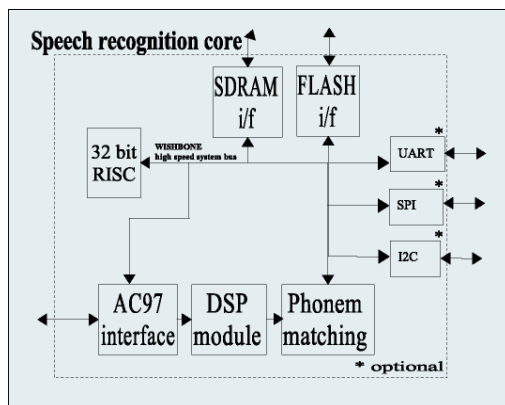
Over the user selectable system interface commands are transferred depending on application and utterances

## Architectural Description

The IP core is made with the following building blocks:

- Analog to Digital Converter (AC97) interface
- Digital Signal Processing (DSP) module
- Phoneme matching
- 32 bit RISC processor core
- SDRAM controller
- FLASH memory interface
- Optional UART, SPI or I2C interfaces
- Wishbone bus

Architecture overview below:



The design uses an external AC '97 version 2.2 compatible audio codec defined by Intel.

The DSP module processes the incoming data stream to extract characteristic data representing the utterance. The data extracted is forwarded to the phoneme-matching module.

A dedicated module matches the extracted characteristic data with the data stored in the phoneme database according to the hypothesis created based on the lexicon and grammar.

The processor dynamically creates hypotheses to be evaluated by the phoneme-matching module. The number of concurrent hypotheses is parameter selected and affects the amount of external memory used by the application. Cost effective high bandwidth external memory is used. First release will support 100 MHz SDRAM.

The Wishbone high performance system bus is used for data flows between internal and external modules, and for interfaces to other systems. Possible solutions include Universal Asynchronous Receiver Transmitter (UART) or serial peripheral buses SPI or I2C. Usage of a peripheral bus makes it easy to include new system interfaces.

The following external resources are needed:

- AC97 Codec
- SDRAM, 1 to 16 Mbytes depending on number of concurrent hypotheses and size of phoneme database
- FLASH, 512 kbytes to 16 Mbytes depending on size of phoneme database.

### Ordering Information

The core can be ordered in the following variants:


- Hard macro, presynthesised netlist for a given technology. Supported target technologies include Xilinx SpartanII and Altera ACEX. Delivery includes behavioral VHDL model.
- Soft core. Synthesisable VHDL core. Script for implementation for different target technologies and synthesis tools included.

For customers building their own speech applications, Voxi provides the Voximizer™ development platform which supports Motorola S-records for easy configuration of FLASH memories.

### Time Plan

The preliminary time plan for the VoxIC™ is as follows:

- First official release is scheduled for Q1 2003. This release will incorporate understanding of predefined utterances.
- Natural language understanding will be added during 2003.

With Voxi's unique approach, previously unachievable solutions are possible. 

### For more information, please contact Voxi:

WWW: [www.voxi.com](http://www.voxi.com)  
 E-mail: [voxic@voxi.com](mailto:voxic@voxi.com)  
 Telephone: +46 8 453 9050