Success story:

Speech recognizer into silicon

Voxi has successfully implemented a speaker independent speech recognizer as an IP core called VoxIC. The design is based on OpenCores IP in combination with specially developed functions for the application. The core is targeted for both FPGA and ASIC devices.

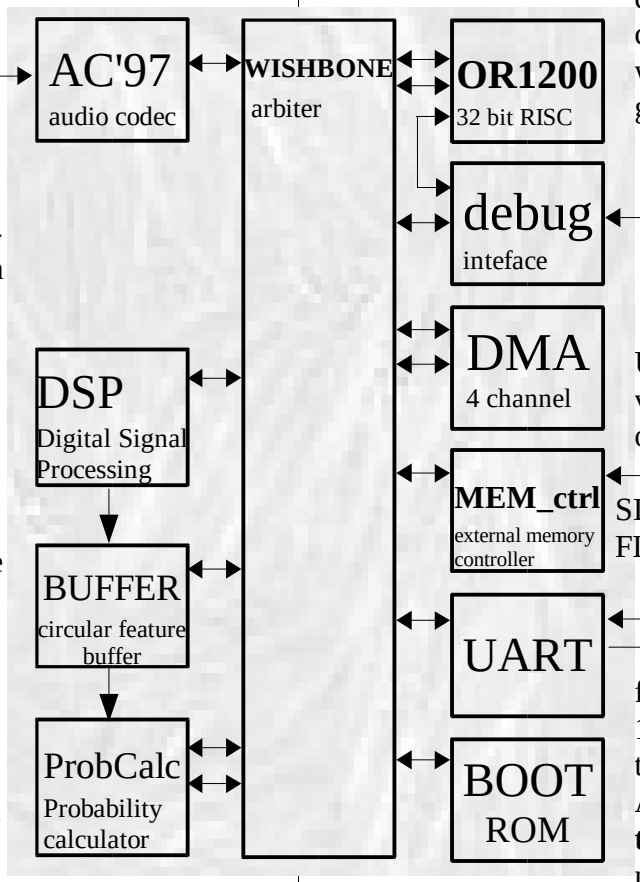
Voxi (http://www.voxi.se) provides the first true **natural speech interaction** between man and machine. Voxi's general-purpose platform for natural speech interaction enables a wide array of applications, ranging from sports to industrial, medicine to inventory control, and entertainment to banking. The object-oriented approach for application design makes natural interaction a standard.

With the completion of the chip integration Voxi is aiming on the consumer electronics market. Implemented as a standalone chip or part of an ASIC the implementation gives a very cost effective speech interface.

The design uses the wishbone system bus. The main reason is the large number of available wishbone compatible IP cores. Since the wishbone bus is a multi-master bus the design can be very flexible. In this design the time critical functions uses master interfaces towards the bus.

As a wishbone arbiter an inhouse developed core is used. Priority of bus cycles is based on master port number. In this way the most time critical tasks are guaranteed bus bandwidth.

Design uses a freely available 32 bit RISC core, OpenRISC 1200, from OpenCores (http://www.opencores.org). The core is highly configurable with optional instruction and data caches, memory management unit, interrupt controller, timer etcetera. The core has powerful debug functions easing software development. The core supports both FPGA and ASIC targets. The main function of the processor is to handle hypothesis of the utterance based on result of processing done in dedicated hardware.



The utterance is processed in two steps:

1. digital signal processing
2. probability calculation

The **digital signal processing** is done in dedicated hardware. All calculations are done with bitserial arithmetic giving a small footprint design. Function JTAG can be provided while the RISC core and external memory are in power down mode. Upon detection of a valid utterance the rest of the design can be woken up. This is important in hand-held battery operated devices. The output from the signal processing is a feature vector every 10:th ms characterising the utterance.

A **probability calculation** is performed to match the feature vector with a phoneme database stored in external memory. To achieve high performance memory access cycles are done in burst mode. All computations are done on the fly.

The flexibility of the wishbone system bus in combination with the availability of IP cores like the OpenRISC has been important factors for the success of this project. The design can be retargeted to any type of device. Additional pheripheral support like compactFLASH or USB can be added to the design with little effort.

/Michael Unneback
Senior hardware designer
Voxi AB