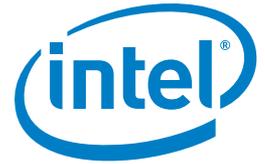


WHITE PAPER

Intel® Atom™ Processor E6xx Series
Nuance* Speech Technologies
Automotive Industry



Delivering In-Vehicle Speech Applications with Computing Headroom to Spare

Nuance* and Intel benchmark the speech performance of an Intel® Atom™ processor-based platform



In-Vehicle Infotainment (IVI) applications, such as digital radio, Internet, DVD video and navigation systems, have evolved from novelties to “must-have” options for many car buyers. Consumers are drawn to cutting-edge features, such as hands free phone calls and voice activated navigation systems, which lead to increased safety and convenience.

This trend was borne out in a recent survey commissioned by Nuance*, a leading provider of speech technologies. In the survey of 473 owners of speech-enabled cars, MAIX Market Research and Consulting Ltd. found a very high usage rate and acceptance level of speech-enabled functions. Eight out of nine participants actively use these functions, with over 70 percent expressing a high level of satisfaction and a willingness to recommend the capabilities to their friends.

When integrating compute-intensive IVI applications, such as speech, developers have to strike the right balance between performance, processing power, system size and cost. To satisfy these requirements, manufacturers of IVI systems need a platform with the performance to deliver innovative features, low power consumption to fit into small spaces, and a high-level of integration to lower cost.

This white paper details the performance of a platform based on the Intel® Atom™ processor E6xx Series running Nuance voice recognition software, which has less than 25 percent CPU processing overhead, leaving ample headroom available to run other applications at the same time. Consequently, the platform eliminates the need for dedicated signal processing hardware for noise enhancement. Particular compiler settings enable the platform to achieve a first response latency of 20 milliseconds for text-to-speech, which is about five times faster than the standard requirement. In addition, the use of Intel® Software Development Products contributed to an over 50 percent signal processing performance improvement.

Boost speech application performance with
Intel® Software Development Products

Speech Use Cases and Technologies

Inside a car is now an environment where many consumers expect an uninterrupted experience of their digital world, complete with navigation, phone connections, high quality entertainment and up-to-the-minute information. Longer commutes and increasingly wired lifestyles are creating strong demand for services that keep drivers connected. They have come to value the ability to control dialing, vehicle functions, in-car entertainment and navigation systems using speech,¹ as illustrated in Figure 1. Some speech functions are performed on-board by the IVI system, and others, like dictating emails or web content, may be performed off-board by servers in the service provider.

Nuance offers speech technologies that facilitate the creation of new services. These technologies were the basis for the benchmark studies that are discussed in the following section. The key components are:

- **Speech recognition:** VoCon* 3200 is a speech recognition engine that accepts natural, conversational input in multiple languages.
- **Speech Synthesis:** Nuance’s Vocalizer for Automotive combines a text-to-speech (TTS) engine, tools and services for enabling speech output tasks.
- **Signal and Noise processing:** VoCon 3200 Speech Signal Enhancement (SSE) removes noise from the microphone input and sends out a filtered signal.

For more information about Nuance automotive solutions, visit <http://www.nuance.com/for-business/by-solution/automotive-products-and-solutions/index.htm>



Intel’s Vision for Enhancing the User Experience

The usability of speech-enabled applications is increasing at a rapid pace, as IVI systems deliver greater accuracy and improved user interfaces. The applications are employing more flexible grammar libraries, which allow drivers to interact using a more conversational dialogue instead of very fixed, predefined menus. Thanks in part to powerful processors, speech applications are accurate and fast, enabling them to be context-aware and respond quicker (i.e., lower latency). Some of today’s processors, such as the Intel Atom processor, have the computing headroom to perform critical noise and echo cancellation functions, thereby eliminating the need for a digital signal processor (DSP) and providing speech applications and far away listeners on phone calls a cleaner input signal.

User interfaces are adapting to the driving environment and are able to integrate better into the multi-modal experience, among other things. With more computing power available in the future, today’s advanced features will soon become commonplace. For instance, natural-language processing (NLP) will expand speech recognition capabilities and enable new services, like a restaurant finder. Systems will automatically adjust the volume of navigation prompts in response to ambient noise, or similarly, lower the radio volume to allow the navigation voice to be heard. Another innovative feature is open microphone speech input that constantly listens for voices so drivers don’t have to physically touch their IVI system to turn it on.

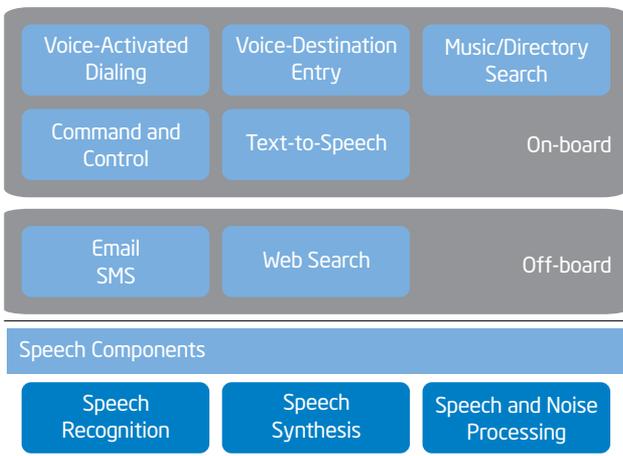


Figure 1. Speech Use Cases and Technologies for In-Vehicle Infotainment

Performance Testing

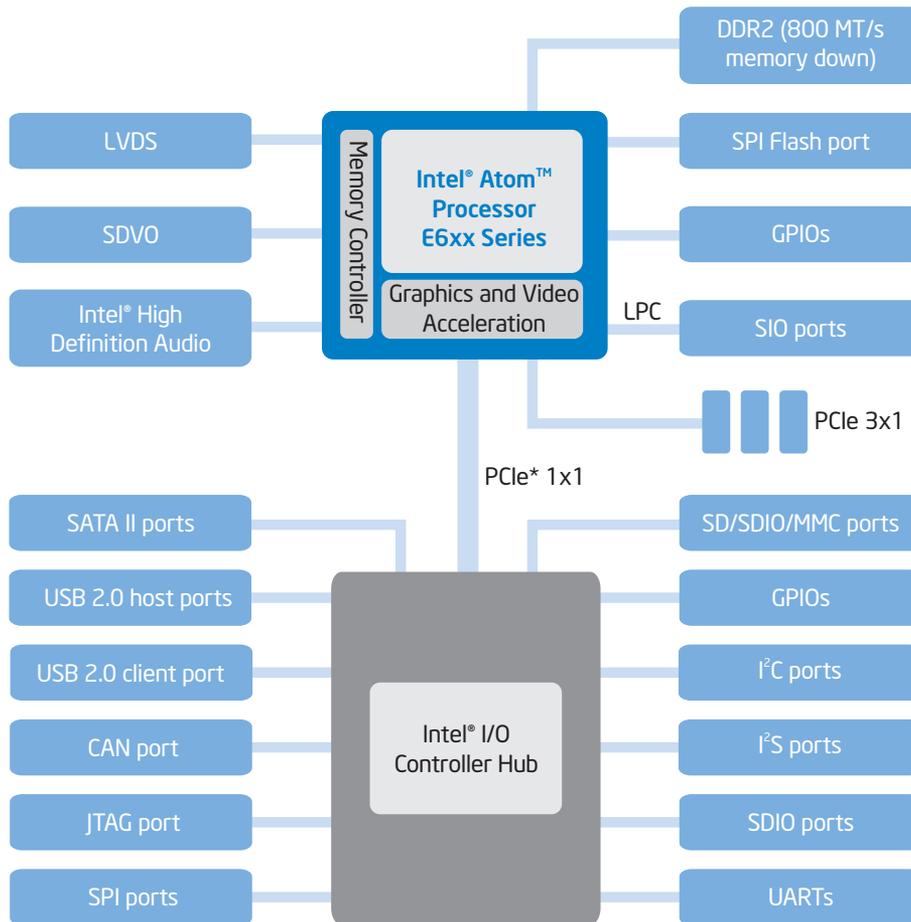
This section provides an overview of the performance characterization conducted by Nuance and Intel on an Intel Atom Processor E6xx Series-based platform, which is briefly described in the following.

Hardware Platform

Enabling a particularly compact and energy-saving IVI system, the Intel Atom processor E6xx series is a very low power system-on-chip (SoC) that offers a new level of platform flexibility. It eliminates proprietary system busses, such as the front-side bus (FSB), and utilizes the open PCI Express* standard for processor-to-chipset interfaces. This allows the processor to be paired with I/O hubs from a variety of vendors that were designed to meet application specific requirements, as illustrated in Figure 2.

The processor is industrial grade (-40° to +85°C), satisfying automotive interior requirements. This highly integrated processor provides for feature-packed computers, such as ultra-sized COM Express* modules measuring 84 × 55 mm or smaller. No fan is needed since the processors in the family consume just 2.7 to 3.9 W (Thermal Design Power - TDP).

Running two software threads simultaneously, the processor performs tasks in parallel, such as navigation and DVD playback applications. At the same time, the on-chip, power-optimized 2D/3D graphics engine enhances visualization applications with minimal load on the processor. Well-suited for visualization and communication tasks, the processor facilitates a multi-modal experience and enables new usage models. Developers can deliver a differentiated user experience and still have computing headroom on-board for future acoustic improvements or to run other applications concurrently.



Note: This figure is a generic representation of the platform.

Figure 2. IVI Platform Based on the Intel® Atom™ Processor E6xx Series

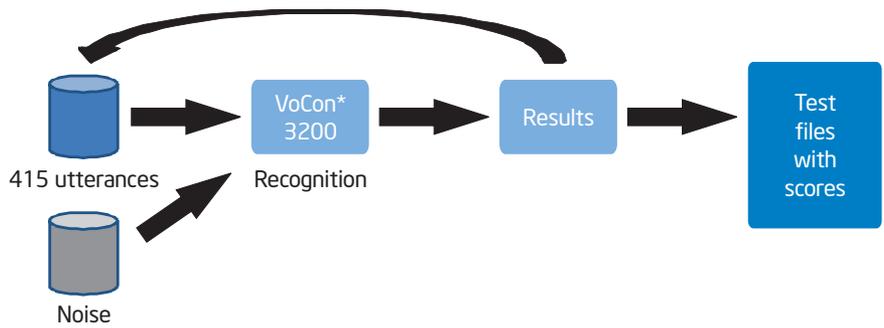


Figure 3. Speech Recognition Test Methodology

Speech Recognition Characterization

Voice destination entry (VDE) is the most resource-intensive of the three on-board speech recognition use cases enabled by VoCon 3200, which also include voice-activated dialing and music search. VDE takes voice input from the driver and maps it across the many ways people express address elements, such as towns, streets, zip codes, house numbers, etc.

Testing was done on 415 pre-recorded utterances spoken by various U.S. English speakers. This workload was issued as a wav file (Figure 3), which was batch processed by the platform. Noise was introduced to assess impact on the system overhead and performance measurements, accomplished by wav file superposition for single channel, including various signal-to-noise ratios (SNRs) to simulate different conditions.

The VoCon 3200 software was compiled and optimized in several different ways. The initial version, or baseline, was a ported version of VoCon 3200 running on the MeeGo* operating system and built with the GCC* compiler. Next, VoCon 3200 was recompiled using the Intel® C++ Compiler with specific optimizations (see Appendix A) for the Intel Atom processor. This resulted in 18.2 percent median improvement in average latency percentage, as shown in Table 1.

Further testing investigated the impact of system memory paging, which was significant given the grammar library was a very large 362 megabyte (MB) file. The default was 20 MB paging, which specifies the maximum amount of system memory that could be used for the library; 20 MB is a recommended setting for many IVI systems available today. The implication is that when VoCon 3200 required data that wasn't already in system memory, a new page would be loaded from the solid state drive (SSD), which added considerable latency.

However, with increased memory capacity available on the Intel Atom processor-based platform, it's possible to employ much larger pages, allowing the entire grammar file to be loaded into memory with paging disabled and thus, improving performance. Demonstrating this performance improvement, another test was run with paging disabled, so there was no set limit on the amount of system memory available to VoCon 3200. This change produced an additional seven percent improvement in average latency, attributable to reduced SSD wait times.

The Nuance software utilized less than 25 percent of the Intel Atom processor, thereby making plenty of computing headroom available to other applications running on an IVI system.³

VoCon* 3200 Performance parameter	Latency Improvement using the Intel® C++ Compiler compared to the GCC* Compiler
Average latency	17.61%
Median latency	18.23%
Maximum latency	18.23%
Maximum latency	26.79%

Table 1. Performance improvements for Latency on VoCon* 3200 Compiled Intel® C++ Compilers as compared to GCC*

Vocalizer Performance Parameter	Performance Improvement: From GCC* Compiler to Intel® C++ Compiler (higher is better)
Cumulative latency	40%
First response latency	34%
Processing overhead	38%

Table 2. Comparison of Average Latency percentage for VoCon* 3200 Compiled with the GCC* and Intel® C++ Compilers

Speech Synthesis Characterization

Further Intel platform characterization was done using the Nuance Vocalizer for Automotive, a text-to-speech (TTS) engine for speech synthesis. The key metric, like the last test, is average first response latency because it is imperative to minimize user wait time.

The Nuance Vocalizer for Automotive was compiled with two compilers, as in the prior VoCon 3200 study. Compared to the GCC compiler, the Intel C++ compiler produced a dramatic 40 percent improvement in cumulative latency and 34 percent improvement in first response latency, as shown in Table 2. In addition, the Intel Atom Processor E6xx Series-based platform demonstrated first response latency as low as 20 milliseconds, which is about five times faster than the standard industry requirement. The processing overhead was also reduced by 38 percent.

Signal and Noise Processing Characterization

The final characterization study was done on the VoCon 3200 Speech Signal Enhancement (SSE) software, which performs signal processing on the microphone input and removes noise. The quality of the output signal can be improved further by increasing the number of microphones in the vehicle. This can be seen by using the Perceptual Evaluation of Speech Quality (PESQ) measurement metrics, which provides an algorithmic means to quantify voice quality relative to actual human listeners (see sidebar).

The benefit of adding microphone channels is shown in Figure 4, where the delta in the PESQ MOS score is measured for different channel configurations. The addition of microphone channels notably improves the sound quality, as denoted by the higher values.

Voice Quality Improvement For Different Numbers of Channels

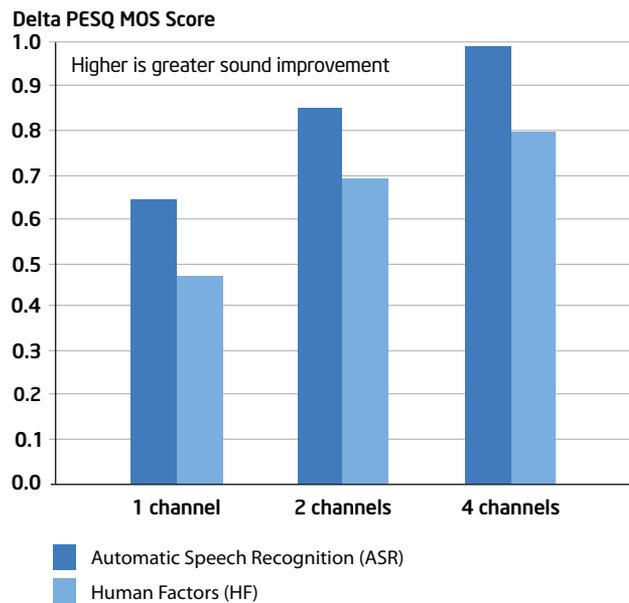


Figure 4. Voice Quality Improvement from Adding Microphones



When using more microphones, the signal processing workload increases for the IVI system. This is illustrated in Figure 5, showing the average normalized latency increasing somewhat linearly with the number of microphone channels. Yet, the Intel Atom processor is capable of processing inputs from multiple sources simultaneously, with low CPU overhead and without using a DSP, which can lower system cost.

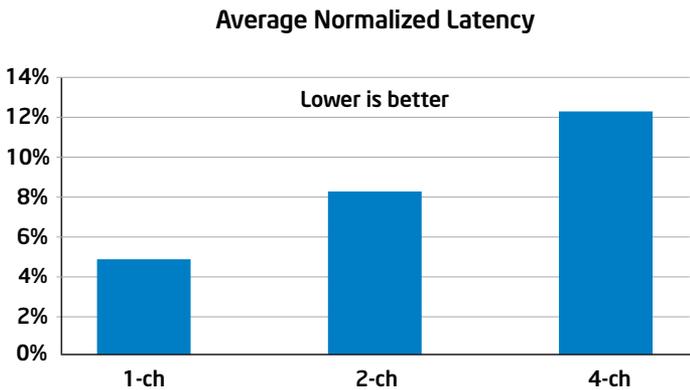


Figure 5. Performance Impact from Adding Microphone Channels

WHAT IS PESQ?

PESQ stands for “Perceptual Evaluation of Speech Quality” and is an enhanced perceptual quality measurement for voice quality in telecommunications according to models of the human perception. Today, PESQ is an international metric for measuring end-to-end voice quality.

The leading subjective measurement of voice quality is the mean opinion score (MOS), based on a large number of people listening to audio and giving their opinion of the call quality, as illustrated in Figure 6. MOS scores, ranging from “very satisfied” to “not recommended,” are mapped to “R” factors, which can be generated electronically and account for network impairments and delays. Combining the best aspects of its predecessors, PESQ is acknowledged for its high degree of correlation to subjective MOS testing.



Figure 6. MOS Diagram

Case	Average Latency Improvement over GCC* Compiler Baseline
Intel® C++ compiler (version 11.1.109) with SIMD-optimization	37.2 %
Intel C++ compiler (version 11.1.109) with SIMD-optimization plus Intel® Integrated Performance Primitives (Intel® IPP) functions	54.9 %

Table 3. Average Latency improvement using SIMD instructions and Intel® IPP

Again, the code was compiled using the GCC and Intel C++ compilers, but this time with single-instruction, multiple-data (SIMD) instructions (see sidebar) and Intel® Integrated Performance Primitives (Intel® IPP), a library of highly optimized routines for the handling of multimedia formats. Table 3 shows the latency reduction (i.e., performance gains) over using the GCC compiler. The results demonstrate that voice enhancement does not require dedicated DSP for enhancement, since the Intel Atom processor can execute the algorithm while maintaining sufficient performance headroom.

Intel® Software Development Products Overview

Developers of signal processing applications have a wide choice of development tools from Intel and the broad Intel ecosystem. The benefits of using these comprehensive tool suites are many, and the tools are applicable to every phase of the software development process.

Intel® C++ Compiler

The Intel C++ Compilers for Linux* and Microsoft* Windows* operating systems are optimized to harness key properties of Intel® architecture processors and deliver optimal performance. They take advantage of a complex set of heuristics to decide which assembly instructions can best optimize the performance in various areas, including memory access, branch prediction, vectorization and floating point operations.

Intel® Integrated Performance Primitives (Intel® IPP)

Intel Integrated Performance Primitives offers a rich set of library functions and codecs capable of speeding up the development of highly optimized routines for the handling of multimedia formats and data of any kind. They have been hand optimized at a low level to provide maximum performance and ease of use with Intel architecture processor-based platforms.

Intel® Parallel Studio XE 2011

Intel® Parallel Studio XE combines Intel's industry-leading C/C++ compilers, performance and parallel libraries, error checking, code robustness and performance profiling tools into a single suite offering. This tool helps developers boost application performance and increase the code quality, security and reliability.

For more information about Intel® Software Development Products, please visit <http://software.intel.com/en-us/intel-sdp-home>.

WHAT ARE SIMD INSTRUCTIONS?

Many signal processing applications are highly parallel, performing the same arithmetic operation on large number sets. Speeding up these workloads, single-instruction, multiple-data (SIMD) instructions were introduced in the mid 1990's, and they perform the same operation on multiple data elements simultaneously, as illustrated below. The throughput of a SIMD instruction is a function of register size because larger registers translate into greater throughput.

SIMD	2	3	5	11	20
+	9	11	2	1	5
=	11	14	7	12	25

The Intel® Atom™ processor E6xx series supports Intel® Streaming SIMD Extensions 2 and 3 (Intel® SSE2 and Intel® SSE3) and Supplemental Streaming SIMD Extensions 3 (SSSE3).

Maximizing Speech Application Performance

Speech applications are requiring more computing performance to increase accuracy and improve user interfaces, and developers can meet these challenges using the Intel Atom processor and Intel Software Development Products chain. The characterization studies presented in the paper identify software development tools and compiler settings that can yield dramatic performance improvements, greater than 50 percent, thus achieving a high return on investment (ROI). In an IVI system, the Intel Atom processor not only delivers exceptional speech performance, but it also has the headroom to run other demanding applications concurrently.

APPENDIX A:

Intel® C++ Compiler Optimizations for the Intel® Atom™ Processor

The Nuance* VoCon* 3200 software was compiled using the Intel C++ Compiler and the following performance optimization switches:

```
> ICC (-O3 -ipo -xSSE3_ATOM -ansi-alias -prof_gen/-prof_use)
```

These switches are not specific to this use case and performance gains may vary depending on the specific application. The (-ipo -prof_gen/-prof-use) enable best possible inter-procedural optimization in the code. The code is instrumented and runtime performance data is collected during a typical execution, which is consumed by the compiler to optimize the final build.

For more information about Intel in-vehicle infotainment solutions, please visit http://www.intel.com/p/en_US/embedded/applications/in-vehicle-infotainment

¹Source: <http://www.nuance.com/industries/automotive/whitepapers/AutomotiveConnectedCarWP.pdf>

²Source: PESQ website at <http://www.pesq.org>

³Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

