



FONIX VOICEGAMING™ SDK



Enhance the fun and power of your game with Voice Control!

New Functionality and Improved Features

- Improved voice recognition accuracy in noisy environments
- New adaptation tool allows for greater recognition accuracy among multiple, diverse and unique dialects, commands and gaming environments, etc.
- Broad language support: US English, UK English, German, European French, Canadian French, Castilian Spanish, Latin American Spanish, Italian, Japanese, Korean and Mandarin Chinese
- Create custom language voice recognition modules
- More detailed recognition information provides greater accuracy and flexibility to developers
- USB and Wii Speak™ microphone support with sample code
- Plug and play ready with the XBOX 360 Kinect™ system

Features and Specifications

Demonstration

To show the many features, recognition quality and the thrill of adding voice recognition to any game, a demo of the Fonix VoiceGaming SDK has been prepared and can be found [here](#). You can add your own words, add different pronunciations, try out the adaptation tool or just have fun with the interactive demo as it responds to your voice commands.

Platform Support

The SDK supports the following platforms: Sony PS3®, Microsoft XBOX 360®, Nintendo Wii™ and PC/Win32. For Apple's iPhone®, iPad® and iTouch® support, please refer to SpeechFX's Fonix iVoicelt™ SDK.

High Accuracy

Using the patented SpeechFX voice recognition process and proprietary neural network architecture, the SDK delivers a speaker-independent voice recognition engine that achieves high accuracy in noisy game environments. SpeechFX software also provides a variety of proprietary tools that allow developers to further improve recognition accuracy for difficult dialects and various background conditions.

Low CPU Usage

The neural network architecture provides the ability for developers to implement complex voice recognition tasks while using only a small portion of the CPU. For example, 5% -10% of the CPU is used for voice processing when the gamer speaks a voice command and significantly less CPU for voice detection in the background. SpeechFX software also provides additional methods to further optimize CPU usage.

Small Memory Footprint

The patented neural network technology is the basis for the voice recognition engine's low memory usage. Total memory (including neural networks and dictionaries) required by the engine is typically less than 500 KB—significantly smaller than competing voice recognition products. The neural network comes in three different sizes, ranging from 50 KB to 250 KB for each language. The smaller size requires less memory and CPU usage but has a small reduction in accuracy. Developers can choose the neural network that best suits their needs based on the trade-off between size, CPU usage and accuracy. Other parameter settings are available to further optimize memory usage.

Sizes: neural networks for US English (other languages are similar in size) are: 87 KB (small), 127 KB (medium), and 231 KB (large). The other two required data files, a user file and vocabulary file, require 150 KB – 250 KB.

Speaker-Independent Recognition

The recognition engine is speaker-independent and does not require user training. The neural networks have been trained on a large amount of speech data representing the varied dialects for any given language and/or country. Accuracy can be further optimized and improved by using SpeechFX's adaptation tool.

Adaptation Tool

The SDK gives developers expanded functionality allowing them to integrate the adaptation technology into a game, giving the gamers the capability of adapting to their own dialects, pronunciations, gaming environments and other commands which are difficult to recognize. For example:

- A command that may sound similar to other commands;
- A command that may have different or alternate pronunciations as compared to a standard pronunciation;
- A game that may be marketed in a particular region where people speak a significantly different dialect from the standard;
- Adverse background gaming conditions.

Developers and/or gamers can apply the adaptation tool to improve accuracy for these cases and more. Increased error reductions from 20% to 50% can be achieved.

User-defined Pronunciation

This tool allows the user to easily change or add to the pronunciations of words. Some words (for example, proper names and places, technical terms, and or newly created words, etc.) may not be found in the standard dictionary. In this case, pronunciations will be matched from the word in the dictionary. The match may not be accurate due to the actual pronunciation not following general pronunciation rules—especially with the proper names and places. Regional and local dialects and/or accents may influence the pronunciation and be different from the standard dictionary. This powerful, new, proprietary feature: “User-defined Pronunciation” tool allows users to modify and self-define pronunciations -- it allows the user to add and define pronunciations that are accurate and correct.

Efficient Memory Management

The SDK has its own proprietary memory manager for efficient memory utilization. It pre-calculates and allocates all required memory during the initialization phase of the game, eliminating the need for allocation during game play. This results in less overhead for voice recognition with no perceptible slow-down or interruption in the game. The SDK also gives developers the flexibility to use their own memory manager if desired.

Flexibility for Game Design

With offline grammar and runtime compilers, the SDK provides flexibility for any game design. For example, a game may need multiple types or levels of speech interfaces and vocabulary groups at different stages in the game. Developers have the ability to build multiple vocabulary files and/or grammars from an unlimited

vocabulary, either during initialization or during game play. In other words, while playing the game, the gamer may select a particular vocabulary based on the current requirements of the game. New vocabularies can then be created during game play to provide further flexibility to the gamer. Using multiple vocabularies (instead of a single vocabulary that includes all commands) reduces CPU and memory usage as well as increases the voice recognition accuracy.

Automatic Speech Detection

Eliminating the need for push-to-talk, SpeechFX's SDK engine can automatically detect the start and end of a voice command. It can also recognize a voice command at the end of speech and then reset itself to wait for the next command. This creates a more direct, interactive experience for the gamer. Pausing for any length of time between spoken commands is also allowed.

Noise Robustness

SpeechFX's neural networks produce highly accurate results in noisy environments. They are specially trained to recognize speech in the presence of many different types of noise, which allows for better voice recognition in noisy conditions. Based on the background noise level and the sensitivity of the audio input device, gamers can also set the noise threshold, a function which triggers the speech detector, in accordance with the level of audio input and the ambient background noise.

Detailed Recognition Results

For each recognition result, the top n-best candidates can be listed (n is set by the developer). Each result provides detailed information which includes the confidence score and the duration for each word in a recognized command (one command may also contain multiple words). Developers can use the detailed information to create a metric differentiation between valid and invalid commands. For example, a valid command usually has a high confidence score and a reasonable duration, while an invalid command may still have a reasonable confidence score, but its duration is significantly shorter than normal speech.

Out of Vocabulary Rejection

Rejection reduces the possibility that the engine will produce a recognition result when you say a word that is not in the vocabulary list. By using the confidence scores for each recognition result, the developer can set a confidence threshold or add a control to allow the gamer to adjust the confidence threshold for accepting recognition results. The threshold level can be adjusted based on factors such as background noise or audio levels, distance of the user from the microphone, etc. The SDK can also provide confidence scores for not only phrases, but also for words. This gives developers more and better information and flexibility to make the most accurate decisions based on the recognized results.

Noise Threshold

With the SDK, developers can design games that give gamers the option of adjusting the noise threshold—the level of audio that triggers recognition. This helps improve voice recognition accuracy in noisy environments. The noise threshold can be adjusted to optimize performance across different headsets, game consoles and ambient noise.

PS3® Spurs SPU Support

For the PS3 platform, the neural networks run on a dedicated Spurs Synergistic Processing Unit (SPU). This offloads a large portion of speech computations from the main processor, which significantly increases the overall efficiency of the voice recognition during game play.

XBOX 360 Kinect

Games that have implemented SpeechFX's Fonix VoiceGaming software are plug and play ready with the XBOX 360 Kinect system. The Kinect audio system houses a multi-array microphone and a built-in echo cancellation module. This allows you to speak commands, in your gaming environment, at distances of 12 feet or greater, which are then processed with the voice recognition engine at amazing split second speeds.

Audio Data Requirements

The SDK is compatible with all standard audio input devices (microphones, headsets) and supports both USB and Wii Speak microphones for the Wii. The neural networks support the following three audio sample rates: 8 KHz, 11 KHz and 16 KHz. If the original sample rate from the audio device is different (for example, the PS3 uses 48 KHz), the data will need to be resampled to one of the supported sample rates. 'Resampling' sample code is included in the SDK.

Grammar Support

Standard BNF grammars are supported by the SDK. Developers can use a simple word spotting grammar to recognize a list of voice commands, or build more sophisticated grammar rules to recognize commands with particular patterns in continuous speech.

Sample Code

Sample code showing how to use the APIs so that developers can quickly implement a powerful voice interface into their games is also included in the SDK.

Language Support

Language support is as follows: US English, UK English, German, European French, Canadian French, Castilian Spanish, Latin American Spanish, Italian, Japanese, Korean and Mandarin Chinese. It also supports Unicode wide characters for inputting and grammars for Japanese, Korean and Chinese. For each language, the SDK provides a dictionary, neural networks with three sizes (small, medium, large) and three sample rates (8 KHz, 11 KHz, 16 KHz).

If a developer needs another language such as Polish, Russian, etc., SpeechFX can build a custom language engine for a fixed set of words (the vocabulary for the specific application). The cost for this service is not included in the standard price of the SDK.

SpeechFX, Inc.

SpeechFX, Inc. ("SFX"), provides embedded, industry leading voice recognition (VR), text-to-speech (TTS) and other technologies to video game developers, creators, publishers and other technology providers which include such companies as: UBISoft®, EA Sports™, THQ®, AutoDesk®, Epic™, Harmonix among others located in the United States, Canada, Asia, Europe and other parts of the world.

SpeechFX Technology in the Game Industry

SFX, a market leader, is a pioneer in developing voice recognition software--specifically customizing it for the needs of game developers and gamers. By providing SDKs and other tools which are optimized for video game operating systems and hardware platforms, developers are able to create an interactive, high performance, voice command and control games across many languages and dialects. SFX technologies provide gamers a proven and more natural interface that allows them to more fully immerse themselves into the game experience through natural voice commands.

