



FONIX VOICEMASTER™ SDK



Enhance language learning skills through instant pronunciation analysis, graphical tutorials and live-audio examples designed to effectively improve pronunciation of sounds, words and sentences.

Fonix VoiceMaster (patent pending) software, developed by SpeechFX, Inc. ("SFX"), delivers powerful, easy to follow tools that greatly enhance language learning methods resulting in a dramatically improved process for mastering a language. It's as if a personal speech professional is standing nearby, teaching correct pronunciation skills. Fonix VoiceMaster, next generation technology available today, is based on standard, international phonetic rules and dictionaries. Designed with extreme utility and flexibility, it allows developers to integrate the Fonix VoiceMaster into various language learning software packages and game applications that can be used on all types of devices such as smartphones, tablets and other small, hand-held devices, as well as laptops, desktops and servers.

In short, here's the software in action. When a problematic pronunciation with a low score is reported, Fonix VoiceMaster provides instant graphical highlights focusing on the particular sounds or phonemes of each single word. With the built-in three intuitive and powerful tools, the user can instantly identify and focus on correcting the pronunciation. To maximize the instructional benefit for the user, these easy to follow tools include graphical, audio and tutorial guides for each phonetic sound. The user continues to correct and improve the mispronunciation by practicing with each tool until accurate pronunciation is achieved. This can all be done at the user's own pace, without the requirement of an instructor to be present.



FONIX VOICEMASTER: THE TECHNOLOGY

Fonix VoiceMaster SDK

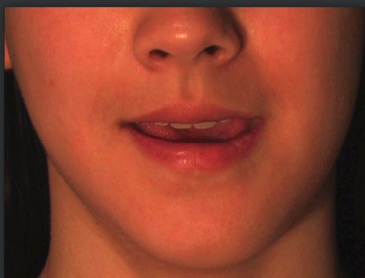
The Fonix VoiceMaster SDK provides software developers with a powerful tool to develop speech-enabled language learning applications. The SDK integrates SFX proprietary voice recognition technology to provide a high-accuracy and low-resource consumption language learning tool. SFX voice recognition software is the foundation for and is widely used by OEMs and other software developers in their consumer products to deliver speech-enabled products, such as voice commands in video games, speech analysis in language learning applications and devices, phoneme recognition for facial animation, etc.

Improving Pronunciation is Easy

Just select a word, phrase or sentence, or add your own word or sentence and click the 'Go' button. Then say the selected word, phrase or sentence and almost instantaneously, an overall numeric score for each word is displayed. Furthermore, each word is divided into separate phonemes with a separate score for each phoneme. These scores are based on IPA (International Phonetic Alphabet), dictionary pronunciation standards and phonetic rules. Red words mean there is an incorrect phoneme pronunciation in the word and red phonemes indicate the phoneme which is not correct. For example, when a sentence is selected, spoken and analyzed, click any word that may be red, then using the step through arrows located underneath the window, the incorrect phoneme can be selected. Follow the prompts and tutoring guides and the intuitive Fonix VoiceMaster tutoring session will assist the user in learning the correct positioning of the articulators of the vocal tract and guide the user to improved pronunciation.

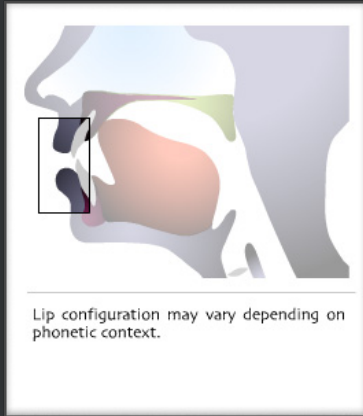
Informative Tutoring Guides

Fonix VoiceMaster contains specially developed tutoring guides that help users understand the placement and position of the mouth, tongue, lips, teeth, ect. (commonly referred to as articulators). This is necessary in order to correctly create the sounds of the target language. In turn, improvement in the pronunciation of words and sentences can be achieved. The three important tutoring guides are:



1- Live Examples

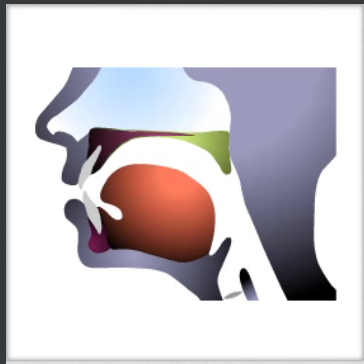
"Live Example" features a live recording of a person saying the sound. A visual example is a strong teaching tool; the user can repeatedly watch how to form and make the sounds while paying close attention to the movement of the various articulators. Two or three practice words that contain the sound, are included. Fonix VoiceMaster re-enforces and confirms the position of the various articulators when combining the target sound with other sounds while forming words and sentences.



Lip configuration may vary depending on phonetic context.

2- Step Through Tutorial

The “Step Through” feature takes a detailed look at the target sound, step by step, with instructions on how to properly position the lips, teeth, tongue, and other articulators in the vocal tract in order to correctly pronounce the target sound being practiced. In this multi-step guide, the graphics show a side view of the face where each physiological area is highlighted with a box describing details of the positions and/or movements required for the making of the particular sounds. A short explanation of each step is also included below the graphics. This allows the user to confirm the positioning of lips, tongue, teeth and other areas of the vocal tract.



3- Animation

The “Animation” feature takes all of the steps from the “Step Through” guide above, combines them together and shows them consecutively as one animated movie clip. This movie clip allows the user to see the various positions and movements of the mouth and articulators when the complete target sound is being pronounced correctly. Sound is also played together in sync with the various parts of the mouth’s and articulators’ movement. This gives the user the capability to try and mimic the target sounds and various positions of the mouth and articulators.

Demonstration

To understand the powerful capability of enhancing any language learning software, whether it be on-line, on a DVD or even at live sessions, check out the demo of the Fonix VoiceMaster SDK [here](#). It includes a pronunciation evaluator along with self-tutoring guides that are designed to improve the sounds of any word or sentence. Try out your pronunciation skills and see how easily you can improve.

SOFTWARE FEATURES AND HIGHLIGHTS

Accent Reduction / Modification

Voice Master is also an excellent tool for carryover and accent reduction/modification. In a therapy setting, as well as at home, the user can work on producing correct sounds, syllables, words, phrases and sentences. This tool provides immediate feedback, and gives the user the opportunity to compare their own production with what is heard and seen from the graphical tutorial guides in the tool.

Phoneme Grading Based on Dictionary

SFX patented neural network voice recognition software allows the uttered word or sentence to be broken into separate phonemes or sounds. The sounds are then compared to a standard dictionary phoneme set and are graded according to how close the sound is to the corresponding phoneme in the dictionary.

Three Pronunciation Practice Structures

There are three input practice structures supplied with the SDK which are: “Words”, “Sentences”, and “Add Your Own”. This means the user can practice on pre-defined words or sentences, or on words and sentences that can be added by the user. For example, the following is one common, practical use. A company’s online or CD language learning instructional workbook may set up practice exercises by allowing the user to click on the words, which in most cases are

organized by chapters in the textbook. The user will select the chapter and the vocabulary list for that chapter. That vocabulary list will then be loaded and made ready for practicing. Likewise, the same would apply for practicing each sentence in each chapter. As mentioned above, the one special unique feature of Fonix VoiceMaster lets the user add any word, phrase or sentence. This capability allows the user to more closely target and focus on specific individual needs making this a very useful and intuitive language learning tool. (For most words and vocabularies, proper names often have non-standard spelling and pronunciations which cannot be guaranteed in this tool.)

Programmable to Automatically Adjust Grading Strictness in Relation to User's Progress

Fonix VoiceMaster includes APIs that give the developer the capability of automatically adjusting the numeric grading system from a beginner, low-level grading system up to an advanced, high-level grading system. This can all be done in real time in accordance to how well the user's pronunciation is. For example, the user practicing pronunciation can set and adjust the initial pronunciation numeric value--low to high, for a particular set of words or sentence activities. Based on the individual's achievement, after the user has reached the initial score for a predetermined number of iterations of a word or sentence, the software would automatically be adjusted to a higher threshold value. This feature, in essence, allows the user to "pass-off" a certain level and increase the performance standard, continually striving to improve pronunciation skills.

Grading Abnormalities

The grading system of neural nets in VoiceMaster is a very intricate and detailed process. It takes into account and extracts many things from the user's speech in order to provide the highest quality of a grading system. In the course of practicing words and sentences, the user should not focus on getting an exact match or the exact same score every time he/she says the word or sentence that is being practiced. No two utterances will ever be the same for a variety of reasons. It could be placement of the tongue, opening of the mouth, rounding or not rounding of the lips, the force of air coming through the mouth, emotions and/or the type of environment the user is in at the time. Background noise or intermittent noise will impact on scoring. The important thing is that scores within the passing range are correct and the words and scores are black, versus scores below the passing range are incorrect and those words and scores are red. Identification of error sounds is the first step in being able to improve speech, so relax and enjoy your tutoring sessions!

Capability of Integrating Custom Dictionaries

Developers of language companies may want their language learning software to be based solely on a specific dictionary and its pronunciation tables. Fonix VoiceMaster has the flexibility to integrate and base its grading and scoring system on the dictionary of choice.

User Defined Pronunciation File

This feature allows developers (may be opened up to users as well) to add pronunciations for words. This can be very useful. For example, (1) if a word cannot be found in the dictionary, its pronunciation is generated by dictionary rules. However, rule-generated pronunciations may not be accurate for all proper names. The developer and/or user will be able to define the accurate pronunciation. (2) Some pronunciations from a dictionary may not be accurate. For example, words may have alternative or different pronunciations in different geographical regions. In this case, developers and/or users can define a pronunciation that is different from the dictionary pronunciation. The input format can be either IPA (International Phonetic Alphabet) symbols or IPA Unicode. The Unicode can be either Unicode format or ASCII format.

Ability to Use Pre-recorded Wave Files

If the developers language learning curriculum or software already contains recorded wave files, these can be accessed and played through VoiceMaster. With VoiceMasters playback feature, the user can replay the models as often as needed in trying to master the desired text.

High Accuracy

The SFX voice recognition process and proprietary neural network architecture provides a speaker independent voice recognition engine that achieves high accuracy in phoneme recognition.

Low CPU Usage

The proprietary 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 - 20 MIPS of the CPU is used for voice processing when the user speaks a word (can vary based on the CPU architecture and number of words spoken). SFX software also provides additional methods to further optimize CPU usage, if necessary.

Small Memory Footprint

The patented neural network technology is the basis for the voice recognition engine's low memory needs. Memory for the base engine is 325 KB ROM, the US English dictionary ranges from 0 - 2,323 KB ROM (dependent upon functionality) and the neural networks come in three different sizes. The neural networks for US English (other languages will be similar in size) are: 87 KB (small), 127 KB (medium) and 231 KB (large). 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. Data size: 200 KB – 1 MB RAM based on number of words or grammar size (approximately 1.2 KB per word ROM/RAM). Audio/Video size: about 78 MB. Total SDK size (US English): under 90 MB. Other parameter settings are available to further optimize memory usage if necessary.

Speaker-Independent Recognition

The recognition engine is speaker-independent, so anyone can use the software. It will grade the user equally, no matter what country or region the user may be from.

Recognition Mode

Characteristics between a child's voice and an adult's voice are prevalent enough to make a difference in the scoring of phonemes. Fonix VoiceMaster takes this into account by supplying a "Child Mode" and an "Adult Mode". The break point between "Child Mode" and "Adult Mode" is about 12 to 14 years old.

Easy to Use

The Fonix VoiceMaster SDK is easy to use. It has a set of clearly defined and simple-to-use APIs with detailed documentation. The SDK also includes sample programs showing how to use the APIs. These features allow developers to quickly build robust, speech-enabled language learning applications.

Environment

The performance of the speech recognizer is robust in noisy environments. The neural networks have been trained with speech data in both quiet and noisy environments. For a more accurate analysis of the user's pronunciation, a quiet or low noise environment is preferred.

Phoneme Extraction

Fonix VoiceMaster has made significant improvements on accuracy and robustness for extracting phoneme information from audio data. The SDK uses SFX proprietary neural network

For more info go to www.speechfxinc.com or call SFX at 801.382.7997

technology to extract the detailed and accurate phoneme information from the audio data. Fonix VoiceMaster uses this information to analyze and evaluate the user's pronunciation.

Phoneme Results

Each result provides single phoneme information for each word as well as every phoneme for each word in a sentence. With this information, users can focus not only on the word(s) that requires better pronunciation, but also on each single phoneme within the word to better improve pronunciation of the individual sounds.

Minimum Score Level

Included in the SDK is the ability to change the Minimum Score Level which is used to judge whether the pronunciation is acceptable or not. If the pronunciation score is above this numeric value, the phoneme and the phoneme score will be black, but if the pronunciation score is below this numeric value, the phoneme and the phoneme score will be red, meaning pronunciation practice is needed.

Audio Data Requirements

The SDK is compatible with all standard audio input devices (microphones, headsets) and also supports USB microphones. 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, the data will need to be resampled to one of the supported sample rates. Audio requirements are as follows:

- 8 KHz – 8 bit μ law/alaw (mono channel)
- 8 KHz, 11 KHz, 16 KHz 10-16 bit linear, mono-channel
- Bandwidth: 100 Hz to sample frequency/2 -- LPF 3dB skirt
- Spectral Flatness: +/- 6dB
- Total Harmonic Distortion (THD) < 2%
- Signal-to-noise (SNR) > -6dB

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.

Language Support

Language support is currently for learning US English. Plans for future languages include Castilian Spanish, Latin American Spanish, Italian, Japanese, Korean, Mandarin Chinese, German, European French, Canadian French, UK English and others. For the Asian languages, support for Unicode wide characters as well as support for grammars is provided. 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).

Important

The grading accuracy of this software, in a large part, depends on the microphone being used, internal computer noise, the audio card, etc. A built-in microphone or a cheap stand-alone microphone may not be a good option as it will not properly shield against computer noise, power supply noise and other types of 'white' noise. This type of 'white' noise may cause degradation of vowels and consonants when the software evaluates and grades your pronunciation. For unaffected scoring and maximum results, please use a good headset microphone or a good quality stand-alone microphone. We suggest doing a recording test with your microphone before beginning. Make sure the audio is clear—without pops, clicks echoes, hums, buzzes, etc. Also, needless to say, the room will need to be fairly quiet unless a good headset microphone is used.

For more info go to www.speechfxinc.com or call SFX at 801.382.7997

Fonix Embedded Programming Interface

This chart shows the OS Commands Pushed into FNX API.

	Win 32/64	QNX	Windows Mobile	Mac OS	Linux 32/64	iOS	Android
C/C++	●	●	●	●	●	●	*
Java	●	●		●	●		*
VB	●						
C#	●		●				

* = Future ports

Platform Support

Fonix VoiceMaster SDK provides support on many platforms (operating systems and processors). This wide platform support provides developers with abilities and flexibility to build applications on a wide spectrum of platforms and devices. MP4 support is necessary for the videos. However, this format can be converted to other compatible formats.

OS Hardware-Platform	Win 32/64	Linux 32/64	OSX Tiger & Lion	Linux 2.2 / 2.4 / 2.6	Windows Mobile	iOS 3, 4, 5	Android 2.0 +	QNX 6.3	Symbian-Nokia Series 60 and UIQ	No OS
<i>Analog Devices</i>										
Blackfin 533/535										●
<i>ARM</i>										
ARM 7		●			●					●
ARM 9/V4/V4i		●		●	●	●	●	●	●	●
ARM 11		●		●	●	●	●	●	●	●
<i>Epson</i>										
S1C33 Family, GNU33										●
<i>Freescale (Motorola)</i>										
PowerPC 5200								●		
<i>MIPS</i>										
R4xxx					●					
<i>Renesas (Hitachi)</i>										
SH3					●					
SH4				●	●			●		
<i>Texas Instruments</i>										
OMAP 1- 5 Platforms				●	●		●	●	●	●
<i>Intel/AMD</i>										
i386+	●	●	●					●		●

SpeechFX, Inc.

SpeechFX, Inc. is an innovative voice recognition (VR) and text-to-speech (TTS) technology company since 1994. It has provided proprietary speech technology to many major electronic manufacturers (Casio, Canon, Seiko, Inventec Besta, etc.) as well as electronic game companies (Sega, Electronic Arts, Ubisoft, Vivendi, Harmonix, A2M, THQ, etc.) and other technology companies in the United States, Asia and Europe.



For more info go to www.speechfxinc.com or call SFX at 801.382.7997