

Speech Recognition

TECHNOLOGY INNOVATIONS APPLIED TO LANGUAGE LEARNING

The advent of computer technologies and multimedia has made language learning easier and more interactive over the past fifteen years.

Ever since it was founded, Auralog has invested heavily in this technological and educational adventure. As a result, it has become the reference for language teaching by CD-ROM and Internet. Thanks to its advanced speech recognition technology, Auralog offers ongoing innovative solutions.

To fully understand the impact of this innovation, one must look back at the evolution of language learning and the techniques associated with it.

How oral methods of language teaching have evolved

Oral methods of teaching foreign languages have seen three major stages of development.

At the very beginning, there was only one way for learners to improve spoken pronunciation: reproduce, as closely as possible, the **teacher's voice**.

Then came the possibility of listening to **recordings** of one's own voice. Early analogue recordings gave way to digital technology, but despite this improvement in sound quality, the technique was still limited by the subject's hearing abilities.

At the beginning of the 90s, Auralog introduced the first method involving **speech recognition**, a technological breakthrough which allowed the student to converse freely with the computer without the help of the keyboard or mouse, and to receive an automatic pronunciation assessment.

What is speech recognition?

The term 'speech recognition' encompasses the entire range of technologies that allow a machine to recognize speech. Modern speech recognition technologies are based on an analysis of the phonemes which make up the spoken sentence.

Auralog was the first company in the world to incorporate this technology into the teaching of foreign languages.

Classifying speech recognition

Three criteria distinguish the various speech recognition engines:

1. The first criterion is the type of speech ("word by word" or "continuous speech") recognized by the software:

- either the words are pronounced **individually**, separated by moments of silence.
- or the speech is **continuous**, allowing for a more natural pronunciation.

2. The second criterion distinguishes speech recognition engines that can process individual **words** only from those that can handle complete **sentences**.

3. The third criterion is the degree of dependence on the speaker:

- A "**speaker-dependent**" system is adapted or adaptable to a given speaker who must familiarize the computer with his or her voice.
- A "**speaker-independent**" system can use speech recognition for a whole group of speakers. This group may include any person speaking the target language (irrespective of age, sex, tone of voice etc...).

Auralog's products use the most sophisticated speech recognition engines, allowing students to perfect the pronunciation of individual words or complete sentences, all at their own pace. What's more, "speaker-independent" software may be used by any speaker and bypasses the familiarization process.

Speech recognition: 25 years of technological evolution

Speech recognition technology has evolved from a hardware stage (cards) into software. Today when a student says a sentence, the software can assess the quality of the pronunciation via a complex analysis of the phonemes.

In the last twenty years speech recognition technology has developed as follows:

Early 1980s

The first **speaker-dependent** speech recognition systems are used by industrial machines.

1985

Around this time the first **cards** equipped with DSP (Digital Signal Processor) appear, enabling speech recognition technology to be installed on a PC. These systems are very difficult to perfect (tests being necessary on a significant number of speakers). These cards are never sold on the retail market.

End of the 80s

Auralog begins adapting speech recognition engines to different markets, before aiming specifically at the market in foreign-language teaching aids.

1991

Auralog launches Aura-lang, the first language-learning application based on speech recognition technology.

1994

DSP cards aimed at the **general public** (using Windows™) reach the shops.

1995

The first software applications based on card technology start to appear. Auralog launches **TaLk to Me**, a retail-market Windows™-compatible software which features a speaker-independent speech recognition system able to process complete sentences. Its speech recognition allows **TaLk to Me** to analyze the student's pronunciation and evaluate it using a sophisticated scoring system.

1996

The use made of speech recognition diversifies into two main camps: language-learning methods and dictation software.

As the market for language software takes off, the first software applications for **dictation** are launched, primarily by DragonSystems. Speech recognition functions on a word-by-word basis.

1998

Auralog launches **TELL ME MORE**, a comprehensive language-learning method which for the first time applies speech recognition to language exercises.

The dictation software market sees the appearance of the first applications allowing **continuous dictation** (The main product is IBM ViaVoice).

1999

Auralog perfects its revolutionary **S.E.T.S.** technology, which is featured in the new version of "**Talk to Me**".

Up until this point, speech recognition had been made to function for continuous speech, on complete sentences and independently of the speaker. The one remaining problem facing development teams was how, within a complete sentence, to identify the word where a pronunciation error had occurred. By locating the weaknesses in the student's pronunciation, **S.E.T.S.** represents a crucial step forward in computer-assisted foreign language teaching. This technology features on the majority of Auralog's publications, notably on the "**Talk to Me**" method.

2000

Auralog offers a language-learning method for English aimed at children aged 5-11, using speech recognition. This is the first time that this technology has been adapted for children.

2001

Auralog launches an online version of TELL ME MORE, making it the first Internet language method with speech-recognition.

2005

Auralog's TELL ME MORE Premium includes a previously unpublished activity using speech recognition: Role Play. The Role Play allows students to approach the spoken language through films and TV films, where they are invited to dub one of the characters. Their pronunciation is evaluated progressively.

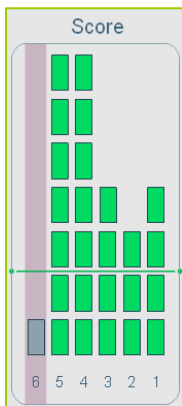
How Auralog's language methods use speech recognition

1. The interactive dialogue



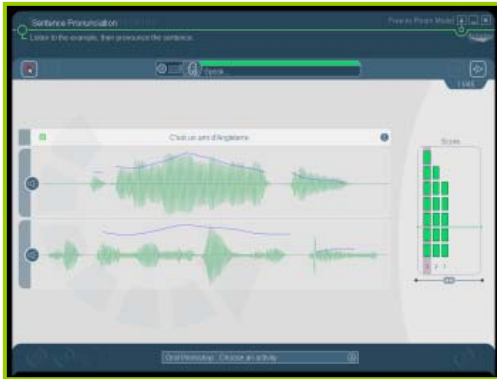
With advanced speech recognition, the student can enter into a genuine dialogue with the computer. The user's responses guide the conversation. The student can adapt the speech recognition to his or her own level, making the computer's evaluation of pronunciation quality more tolerant or more demanding.

2. Pronouncing words and sentences

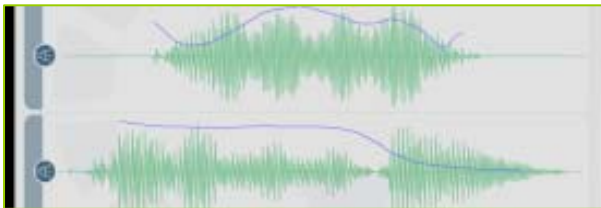


The user practices pronouncing a word or sentence, and receives a score which can be used to evaluate the quality of accent, pronunciation and intonation.

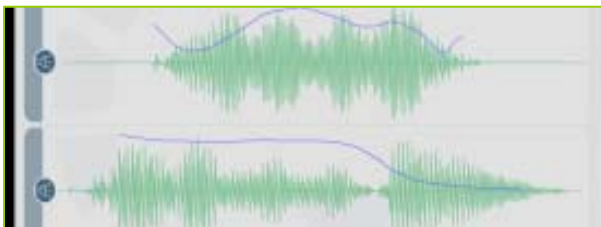
3. How pronunciation is displayed



As well as the scoring system, the software also allows the user to visualize accurately not only pronunciation but also *intonation*. Two types of display mode (waveform and pitch curve) are provided. The user can display them at the same time, or individually.



The waveform indicates the voice's volume as a function of time (the notion of energy). It represents the sound intensity of the voice and gives a view of the structure of the pronunciation.



The pitch curve reproduces variations in the voice frequency against time. In tandem with the waveform, this curve enables students to make precise comparisons of their own intonation with that of the model (high-pitched/deep). This unique display mode is an innovation developed by Auralog.

S.E.T.S. - Spoken Error Tracking System



S.E.T.S is a technology that enables the computer to automatically detect where a pronunciation error has occurred. This exclusive technology, dubbed **S.E.T.S. (Spoken Error Tracking System)**, is currently patent pending.

Auralog is the only software publisher to offer applications which evaluate the **pronunciation** and **intonation** both of **complete sentences** and of words and which allows them to be visualized.

4. Role Play

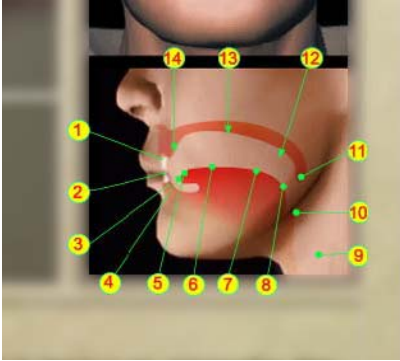


Students can approach the spoken language in a recreational way with TELL ME MORE's Role Play activity. During the dubbing stage, students set the playing speed in order to give themselves time to record their lines.



The Role Play pronunciation workshop offers students the chance to have their lines played back in context, to work on the pronunciation of these lines and to receive an evaluation.

5. 3D animated phonemes



Computer animations of individual phonemes allow the student to observe articulations in close detail, and so reproduce them more accurately.

This exceptional technology helps the student understand the articulations involved in producing certain sounds. This in turn facilitates recall, and leads to better reproduction.

- | | |
|--------------------------|------------------------------|
| 1 - Upper incisors | 8 - Back of the tongue |
| 2 - Lower incisors | 9 - Vocal chords |
| 3 - Alveolar ridge | 10 - Glottis |
| 4 - Tip of the tongue | 11 - Uvula |
| 5 - Blade of the tongue | 12 - Soft palate |
| 6 - Front of the tongue | 13 - Hard palate /
Palate |
| 7 - Centre of the tongue | 14 - Alveolar ridge |

Glossary

Speech recognition

General term for those technologies that allow a machine to recognize speech.

Speech recognition difficulty level

Level at which or above which the computer considers the pronunciation of a word or a sentence to be acceptable. Auralog's products allow the user to set this level, in order to modify the difficulty of the pronunciation exercise.

S.E.T.S. (Spoken Error Tracking System)

A technology exclusively developed by Auralog, which automatically detects pronunciation errors within a complete sentence.

Waveform

Graphic display of the changes in volume of the voice, against time. This graph gives information as to the structure of the pronunciation.

Pitch curve

Graphic display of the variations in voice frequency, against time. This helps to improve intonation.

Continuous speech recognition

A characteristic of the most advanced speech recognition engines. Dialogue with the computer is fluid and does not require stops.

Word-by-word speech recognition

Recognition by the computer words to be pronounced individually, i.e. separated by periods of silence.

Speaker-dependent system

Speech recognition system adapted (or adaptable) to a given speaker, who must familiarize the computer with his or her voice.

Speaker-independent system

Speech recognition system able to work with a whole group of speakers, which is theoretically capable of including any person speaking the target language.

DSP (Digital Signal Processor)

Processor which allows a PC to be fitted with speech recognition hardware. Such systems have now been replaced by speech recognition software.