



SOFTWARE

The computer's voice

Linguists and engineers from Unicamp formulate a speech system with a Brazilian accent

TÂNIA MARQUES

Published in May 2003

If a good number of machines now “speaks” sufficiently well for carrying out simple tasks, and many people have for some years been “talking” to automated telephone answering systems and automatic teller machines, the synthetic voice resources in commercial use still show some difficulties in reproducing human speech with naturality. And their vocabulary is very limited. But there are indications that computers will soon be losing their digital accent and expand their linguistic universe. Big companies are starting to get results that are more natural and agreeable to the ears. This quest for perfection in sounds from computers began early at the State University of Campinas (Unicamp).

A joint project, started in 1991, between the linguistic and electrical engineering areas produced software that today is capable of reading aloud any text written in Portuguese, without the characteristic English accent of the systems produced outside Brazil. The Brazilian program bears the name of Aiuruetê, which means “true parrot” in the Tupy (the most common Brazilian indigenous language).

Right from the beginning, the development of the system has been subordinated to scientific ends, but the project has also produced some technological results. “We wanted to create a speech synthesis system in Brazilian Portuguese, starting with basic research and focused on it”, recalls Professor Eleonora Cavalcante Al-

bano, from the Phonetics and Psycholinguistic Laboratory of the Language Studies Institute (Lafape/IEL), who is coordinating the work. Maintaining the original target and with a broad vision of the phonetic-acoustic description of language, the venture included studies of problems of articulatory development and disturbances, phonological theory, phonostylistics and the analysis and synthesis of speech.

Swift evolution - In 1992, Professor Fábio Violaro, the coordinator of the Digital Speech Processing Laboratory of the Faculty of Electrical Engineering (LPDF/Fec) and his group of researchers embraced Lafape's project. "We were already working with speech synthesis, but the results of our efforts were limited, precisely for the lack of linguistic knowledge", says Violaro. At the time, personal computers were evolving apace, and their resources for processing and memory were already making it possible to develop voice synthesis programs. Today, Aiuruetê runs on any computer with a Windows operating system.

Speech synthesis programs, which can make a big contribution to distance learning and to the education of the visually impaired, besides a series of commercial applications, are usually based on the conversion from text to speech. Like similar foreign software, Aiuruetê works with textual information, which, in the preprocessing stage, is submitted to an analysis, to include the grammatical characteristics (acronyms, abbreviations and graphic symbols) and rewritten in full in the way it is read. Afterwards, it undergoes a phonetic transcription. Then the software looks in its database for utterances compatible with the transcribed material and takes care of stringing together the phonetic elements that make up the words, also giving them information on the intonation and rhythm of Brazilian Portuguese. Does it seem easy? Well, it isn't – so much so that since the beginning of the so-called digital ages speech synthesis has been a challenge to researchers from all over the world, who have attained a level that is no more than reasonable.

Several factors contribute towards the complexity of the process, in any language. In the first place, systems written for different languages have varied degrees of phoneticity – only up to a cer-

tain point does the spelling of words determine their pronunciation. English, for example, has an orthography that is far from being phonetic. Words spelt in a different way, such as rite, write, right and wright are pronounced exactly the same way and have, therefore, the same phonetic transcription: RIT. The orthography of Portuguese has medium phoneticity, but even so does not offer fewer difficulties. To stay with just one example, suffice it to remember that the letter "x" may have the sound of "sh", "s", "ks" or "z". "Portuguese is nice, but Spanish is much better", Eleonora jokes.

Addressing the question, a layman can imagine that the construction of a database with all the words of the language is the solution. But an enterprise of this kind, besides being monumental, would be fated to failure: language is dynamic, and new words arise every day. Furthermore, the pronunciation of one and the same word varies in accordance with the context, which would imply the need for recording the same word several times – there simply could not be any dictionary of such a size. Even words that are widely used may not be in any dictionary, as well as the verbal inflections and the diminutive and superlative forms. What software chiefly needs is parameters to guide the pronunciation by the machine.

"We opted for limiting ourselves to some 2,500 excerpts from recordings", says Eleonora. The number is not a very high one, but the excerpts were submitted to a strict selection. In it, the researchers did not work with a

traditional concept in linguistics that defines the phoneme as the smallest mental unit corresponding to sound. Since the start of the work, the team has maintained the theoretical position according to which the phoneme is an abstraction influenced by alphabetical writing. One of the points of the study was of the various phonemes that undergo an influence from those that precede them and from those that follow them. "Many factors are combined in the articulation of sounds, and a 'p' followed by an 'a' is pronounced differently from 'p' followed by an 'i' or a 'u'", Eleonora observes.

Another problem in developing a speech system is the differences between the graphic representations of the text and the way they are expressed in speech. Abbreviations, for example, can be read differently, even when they have the same number of characters and are equally pronounceable. In this regard, it is worth comparing USA with NASA, for example. Reading a telephone number is different from a numerical expression – nobody would read 32220000 as 32 million, two hundred and twenty thousand. In Portuguese, measurements of length are written in the same way in the singular and in the plural: 1 meter and 100 m. All this calls for complex algorithms.

Emotion and subtleties - "Although it can already be used in a series of applications, Aiuruetê is still under development", explains Violaro. Among the improvements, there is the assimilation of the subtleties of the rhythms of Brazilian speech. "In future, we want Aiuruetê to express even the tonal differentials of emotion", says Eleonora. According to Violaro, the program is beginning to arouse the interest of some companies that are specialized in information technology. One of them was to use Aiuruetê in a self-service system aimed at medical clinics, with the booking of appointments and other functional features. Furthermore, the work will also result in building up a public database of knowledge of the phonic aspects of the Portuguese spoken in Brazil. The software is therefore getting closer to one of the most appreciated properties of true parrots: being the most talkative of the Psittacidae family. •

THE PROJECT

Processing Text and Acoustic Signals in Brazilian Portuguese: A Linguistic – Engineering Interface for the Science and Technology of Speech

MODALITY

Thematic project

COORDINATOR

ELEONORA CAVALCANTE ALBANO –
Language Studies Institute at
Unicamp

INVESTMENT

R\$ 9,528.00 and US\$ 58,672.00