Speech Synthesisers for Atari and Apple

dropbox

Review:

Speech Synthesizers For Atari And Apple
Charles Brannon, Editorial Assistant Tom R. Halfhill, Features Editor

Let's be honest. How many of us have watched Star Trek on TV without wishing that our computers could talk, too?
Synthesized speech has been around for a few years now, but has cost hundreds of dollars, even on microcomputers. That's why two new speech products for the Atari and Apple have stirred so much interest - their quality sets a new high, their prices a new low.
The Alien Group's new Voice Box ranges from $139 to $215, while Don't Ask Computer Software's revolutionary synthesizer on a disk -Software Automatic Mouth (S.A.M.) - checks in for e~n less at $59.95 (Atari version) and $124.95 .)Apple version). Both are capable of startlingly human-like speech.
The two products approach the problems of speech synthesis in quite different ways, however. The Voice Box is a plug-in "little black box" supported by machine language programs that allow you to create and store dictionaries of frequently used words. S.A.M., however, is entirely software-based, using no hardware at all (except for a simple digital-to-analog converter and amplifier board in the Apple version).
Since both products hit the market at almost the same instant, and since both are for two of the most popular personal computer systems, there's bound to be brisk competition as people line up on each side of the which-is-best fence. Therefore, we'll state up front that neither will be declared the clear-cut victor here; both are good products, and each has its strengths and weaknesses.
Keeping that in mind, we can explore several criteria for evaluating microcomputer-based speech synthesizers. These include speech quality, versatility, and the ease of incorporating speech into user-written programs.

Is It Human?
Speech quality is probably the most important of these. How closely does the synthesized speech simulate human speech? Both S.A.M. and the Voice Box speak in recognizable tones which approach human speech very closely. Both voices are male, not because the programmers were sexist, but because female voices are harder to synthesize due to their wider dynamic range.
S.A.M. speaks with a definite accent, although the nationality is hard to place. To some it sounds somewhat Scandinavian, perhaps Swedish. Then again, it might be East European. At any rate, S.A.M. speaks English as if it were a second language. This is not intended as criticism; on the contrary, S.A.M. talks very brightly, enunciating words and syllables with a sense for inflection and accent that is quite amusing. Some syllables sound sort of thick or fuzzy (especially a "th"), as S.A.M. struggles to do with silicon chips what a person does with a tongue and palate.
The Voice Box is distinctly different. It speaks in a smoother voice than S.A.M., without as many fuzzy syllables, although it, too, has trouble with certain sounds (a "g" resembles a "d"). However, the Voice Box tends to speak in a monotone when converting plain English to speech, while S.A.M. adds its own unique intonation. If the Voice Box speaks with any accent at all, it is "computerese": neutral, unemotional. The nuances are hard to describe, but the results are that the Voice Box tends to offer the more human-like tones, while S.A.M. tends toward more human-like speech patterns.

To put this another way, if you were to have each synthesizer read a plain English sentence over the telephone to a person who was unaware that a computer was speaking, the Voice Box would be quickly identified as a computer, while S.A.M. might more easily pass as a human, albeit one with a heavy foreign accent.
Remember, though, we're talking about each product's ability to interpret plain English. English is a formidable challenge because it is a language of as many exceptions as rules. To program a computer with a complete knowledge of English pronunciation - to distinguish between though, bough, and tough, for example, would require massive amounts of time and memory.
Considering this difficulty, S.A.M.'s text-to-speech "Reciter" program works surprisingly well. Given ordinary English text, the Reciter will pronounce it, even adding inflection automatically. The Voice Box uses a "dictionary" to memorize words you "teach" it. If it learns many common patterns such as "ch", "ou", etc., it can simulate a simple text-to-speech algorithm. The advantage of such a dictionary is that you can be sure it will pronounce a memorized word correctly.

The Atoms of English

Much higher quality speech is attainable by using phonemes. Phonemes are the "atoms of English," as S.A.M.'s manual puts it - the basic sounds upon which all spoken words in the language are based. There are only about 50 or 60 of these.
Both products allow you to define words using special combinations of letters, numbers, and symbols representing these phonemes. For instance, S.A.M.'s Reciter has a little trouble pronouncing the word "synthesizer." A much more accurate result can be obtained by leaving the Reciter pro~ram and entering the word as a series of phonemes: "SIH4NTHAXSAYZER."
The Voice Box uses a similar set of phonemes. An example for the same word would be "SI2N-TH-ES-UH3-AH2-Y-ZER." Hyphens are used with the Voice Box to separate the phonemes. To add inflection to words and syllables, you use slashes
- a forward slash (I) raises the pitch and a backward
slash (\) lowers it.
Yes, the phonemes look like alphabet soup, but you must use them for tricky English words if you want accurate speech. Each product lets you vary the pitch, speed, and inflection of speech in enough ways so that virtually any English word is pronounceable. Again, S.A.M. does this entirely with software, while the Voice Box has an additional tuning knob which lets you adjust the overall speed and pitch of the speech from slow and guttural to fast and squeaky, very much like changing the
speed of a tape recorder.
In addition to pitch control, S.A.M. also lets you vary overall speed, and independently stress words or syllables with eight levels of emphasis. Such phoneme-based text is hard to program and read, but it produces some incredibly high-quality speech.
The Voice Box's ten pages of documentation include a phoneme list with example words. S.A.M.'s 40-page manual has a very helpful 15-page dictionary of common words and their phoneme equivalents pre-defined for you.

Programs That Talk

Both products allow you to incorporate speech into your own BASIC language programs. You can now have talking aliens, game instructions, audible error messages, and practically anything else you can think of.
Both synthesizers require that their machine language programs be loaded along with your BASIC program and called as subroutines. The text to be spoken is contained in a string variable. Software included with the Voice Box provides a
"skeleton" program, complete with the machine language necessary to use the "black box," that you can add to your own program. Alternatively, you could start with the framework program and build your application around it.
S.A.M. "boots" (automatically loads) from a copy-protected diskette. It is simpler to interface with your BASIC program, requiring only one setup statement, and two statements to "call" S.A.M Remember, however, that you must always load the actual S.A.M. synthesizer from the special disk. The text-to-speech ~eciter program isjust as simple to use, but must be accessed from a separate disk you prepare. And since S.A.M. is all software, it consumes much more user memory than the Voice Box.
The Atari version of S.A.M. blanks out the screen as it speaks, precluding the possibility of synchronizing speech with graphics. However, the original screen image always returns when S.A.M. has finished. The Voice Box does not blank the screen, but the software which drives it waits until the speech is done, causing a similar freeze while the box is talking. This can be circumvented with tricky machine language, and documentation is provided to help advanced users access the Voice Box from the machine language level. There also is a way to stop S.A.M. from blanking its screen, using a simple POKE, but the result is extremely distorted speech that is impractical for most applications.

Synthelic Shakespeare

Aside from the machine language driver programs, both products supply various uitlities and demos. S.A.M.

provides a guess-the-number game, a simple talker program, and a set of four famous speeches - from the Gettysburg Address to Hamlet's soliloquy.
The disk or cassette supplied with the Voice Box includes the aforementioned skeleton program; a "help" demo that shows how to program accurate speech; a "talking head" that lip-syncs with the voice; and two versions of a talker program for 16K or 32K RAM machines. The extended 32K version includes a random sentence generator which utters outrageous phrases, not unlike some of the stream-of-consciousness poetry popular in the SOs and 60s. An example: "That desk quickly loves your rabbit if a ham sandwich sits on my big small girl when yo~ur rabbit sleeps."
The Voice Box is (at the moment) the only product usable on cassette-based systems, with abridged support software available on cassette.'

A singing Computer?
Although untested, a singing version of the Voice Box is available for the Apple at a higher price.