Review of Three Software Systems for Speech Analysis: CSpeech, BLISS, and CSRE

Évaluation de trois logiciels pour l'analyse de la parole: CSpeech, BLISS et CSRE

John Ryalls
Ecole d'orthophonie et d'audiologie
Université de Montréal

Shari Baum
School of Human Communication Disorders
McGill University

Abstract

Three software packages for digital signal analysis of speech were evaluated and compared. Factors such as user-friendliness and ease of execution relevant to operations likely to be of interest to Speech-language pathology were considered and discussed. Each of these three relatively low-cost University-developed systems runs on the IBM-compatible AT type machine, using the Data Translation DT-2801-A analog/digital conversion card. All three provide 12-bit resolution at the 10 or 20 kHz sampling rates, standard to most speech research.

Résumé

Trois logiciels pour le traitement numérique de la parole ont été évalués et comparés. Des facteurs pertinents à l'orthophonie, telles la convivialité et la simplicité de l'exécution, ont été pris en considération. Chacun de ces trois logiciels a été conçu pour le micro-ordinateur de type AT IBM-compatible, utilisant la carte analogue/numérique de Data Translation. Tous les trois fournissent une résolution de 12 bit pour les taux d'échantillonnage de 10 ou de 20 kHz (standards pour la majorité des recherches sur la parole).

Introduction

In the last decade there has been increasing use of the microcomputer in the speech-language pathology clinic. Digital computer technology almost has usurped the Sonagraph for modern speech research. Once only available in large university research settings, microcomputer-based speech analysis is now within the financial realm of more and more speech-language pathology clinics, because of the decreasing hardware costs combined with the increased computing power and software advances. This is more true because computers often are already available in such clinics for keeping patient records and word-processing. Computer-based speech analysis systems allow a precise characterization of speech production deficits exhibited by a variety of clinical populations. Rather than relying solely on the clinician's perceptions, in addition to these, specific acoustic analyses may yield further insights into the nature of articulation, phonation, and prosodic errors. Speech analysis software thus is not only a useful research tool, but also an important clinical tool. For instance, speech measures may be useful in determining the degree of progress in clinical intervention.

One problem we have had to deal with in setting up our own respective speech research laboratories has been the lack of information available about speech research software packages. This is true particularly of university-based software because its non-commercial use typically does not permit developers to advertise their product to the general market like their commercial equivalents. This is somewhat unfortunate in that these university-based efforts are more likely to meet the needs of individuals conducting speech science research. It is in the interest of increasing information relevant to speech science research and of reducing for others some of the frustrations we felt in getting our laboratories up and running that we offer this evaluative report. Our review intentionally assumes a non-technical position. A somewhat more technical description of two of the systems considered here (i.e. CSpeech and CSRE) has recently become available in Read, Buder and Kent (1990). Technical details can also be obtained from individual program developers.

The three packages we will consider here are: (1) Paul Milenkovic's CSpeech package developed at the University of Wisconsin; (2) John Mertus' BLISS package developed at Brown University; and (3) Donald Jamieson and Terrance Nearey's CSRE (Canadian Speech Research Environment) developed at the Universities of Alberta and Western Ontario. First we will consider the types of analyses available in all three packages and the subprograms included in each. All three programs obviously provide a means of analog-to-digital conversion, or signal digitization, at a number of sampling rates which may be set by the user. In addition, each program includes a waveform editor and a series of routines for spec-

tral analysis. Two of the three programs (BLISS and CSRE) provide for the synthesis of speech from numerical parameters, while all three include subroutines for pitch extraction. In addition, BLISS incorporates a program for the generation of auditory perception and reaction time experiments, including a Grader program for treating experimental data.

In our brief review, we will focus on several acoustic parameters commonly measured in speech analysis, such as duration, one or two spectral analyses, and fundamental frequency extraction, highlighting those aspects of each software package that we have found particularly advantageous or particularly troublesome. Our evaluation is based on laboratory trials in the case of CSRE and CSpeech, while our experience also includes current research projects in the case of BLISS.

CSpeech

CSpeech (for "see" speech, not for "C" language) is, in its developer's description, "a multi-channel program, intended for simultaneous measurement of the acoustic speech signal as well as signals derived from instruments such as the electroglottograph, neck wall accelerometer, air flow mask and intraoral air pressure sensor" (Milenkovic, 1989). The complete CSpeech software package (version 3.0) is available from Paul Milenkovic, its author, for $1800 (U.S.). We found CSpeech to provide a user-friendly work environment. Menus constantly guide the user through the choices available at each point in the program. Sampling is straightforward in CSpeech with the waveform automatically displayed immediately after digitization. One must remember to save the sampled speech at this juncture because re-sampling will write over this file if it is not stored. This is a distinct advantage when one wants to erase what one has just sampled, less so otherwise.

Duration measures may be performed easily by placing a single set of cursors on the digitized waveform; the duration between the cursors is automatically displayed. With regard to spectral measures, CSpeech provides an LPC (Linear Predictive Coding) algorithm that computes an analysis, displays the LPC spectrum, and allows (via cursor movement) precise measures of any point along the LPC contour. The resulting spectrum plots have 10 dB and 1 kHz tick marks. Preemphasis and number of coefficients are user-defined.

We liked the ease with which we could perform LPC analysis but would prefer to be able to set a fixed window over some portion of the speech segment instead of having the window-size vary with the portion of speech between the two cursors. Once the LPC is displayed, one has to scroll through the spectrum to read off formant values. There is no way that the peak values can be printed out with the transfer function. We found this scrolling to be a little time consuming; however, we were happy to have LPC spectra so easily. It is also a distinct plus to be able to print these spectra on ordinary dot-matrix printers that are commonly used with microcomputers. Multiple spectra can be displayed on a single plot.

The manual was far from clear on the pitch analysis procedure, and one is easily confused by the practically identical screens for each channel. In the resulting plot, no frequency scale is provided, thus one must scroll the cursor through the pitch contour in order to read off frequency values. However, it should be pointed out that CSpeech does allow one to perform inverse filtering (estimate of source characteristics) as well as jitter (frequency variation) and shimmer (amplitude variation) analyses, in addition to traditional extraction of fundamental frequency. Clinicians working in voice disorders will find these analyses useful. CSpeech, as noted earlier, does not have a speech synthesis facility. Overall we rated CSpeech very highly in terms of user-friendliness and speech research power.

BLISS

BLISS was developed by John Mertus at Brown University and is available at a cost of $250 (U.S.). It has been described as follows: "BLISS, which stands for Barus Lab Interactive Speech System, is an interactive package of programs for analyzing and processing speech or other digitized signals" (Mertus, 1989). It should be pointed out that the AT version of BLISS has only recently become available and that it is the offspring of Digital's PDP 11-34 and microVAX versions. Its compatibility with a larger frame computer will give it an instant appeal to laboratories that have PC workstations in network with such larger computers. We should mention, however, that further development of the AT version of BLISS is unlikely.

Although sampling (of up to two channels) in BLISS is slightly less straightforward than in either of the other two programs, the program allows better protection against clipping and more user-defined parameters. Of the three software packages, we preferred BLISS's waveform editor.

As in CSpeech, duration measures are very straightforward in BLISS. Four sets of cursors are available for editing and measurement of temporal parameters. Movement from one file position to another is accomplished in one of two ways: gross changes may be made via a simple mouse click, and the program does include a somewhat crude waveform scroll for fine adjustments in cursor placement. In addition, cursors and windows may be moved from within a second subprogram (WAVED) using mathematical expressions. We

missed the fact that the point at which the cursor makes contact with the waveform does not change colour to more easily pinpoint locations, as it does in CSpeech. However, we found the waveform editing screen to be exceptionally powerful and user-oriented.

For spectral analyses, BLISS provides LPC spectra and bark scale plots (i.e., a log scale more reflective of frequency representation along the basilar membrane). The peak values from the LPC analyses may be printed (frequency, bandwidth, and amplitude) or stored for statistical manipulation. As with CSpeech, multiple spectra may be overlaid one on another; but in contrast to CSpeech, axes are labelled in BLISS, and many parameters may be adjusted by the user as needed. On-line help files are available for most applications.

Pitch analysis is more straightforward than in CSpeech, and again, most parameters may be user-defined. BLISS provides jitter measures, but not shimmer measures.

As noted, BLISS also provides subroutines for the generation of auditory perception and reaction-time experiments. The types of experimental paradigms which may be supported include: dichotic listening, lexical decision, and word recognition.

Unfortunately, BLISS's documentation could benefit from a little more completeness and detail. It should be mentioned that BLISS is the only package of the three for which a math co-processor is not required to function. This may be a significant consideration if one is thinking about implementing such software on a laptop computer. BLISS will play back files without the A/D card using the computer's internal speaker, which allows analyses to be performed on previously sampled data. Obviously, the A/D card is required for sampling.

CSRE

While CSRE is not a direct sibling of BLISS, it is at least its Canadian cousin. This program is described in the following manner: "CSRE 3.0 is a comprehensive, integrated, inexpensive, microcomputer-based workstation system to support speech research. The analysis portion of the package contains a set of three waveform editors and a comprehensive facility for spectral analysis" (Jamieson, 1989). Additional information on CSRE can be found in Jamieson, Nearey and Ramji (1989). CSRE 3.0 is available at a cost of $350 (U.S.). According to the author of BLISS, John Mertus, BLISS and CSRE are not competitive projects; rather, they reflect simultaneous developments in the two laboratories. Jamieson and Nearey purchased the A/D and D/A drivers from Mertus; thus audio routines in BLISS and CSRE systems are identical for the Data Translation A/D boards. Eventually, CSRE and BLISS are expected to have similar file structures for total compatibility.

Sampling in CSRE is quite simple, and an immediate display of the digitized signal is provided. For duration measures, however, we found the waveform editor to suffer from accuracy overkill. Although one can position a cursor to within point accuracy, we did not like having to position cursors at three levels of detail in order to obtain simple duration measures. Furthermore, although there are some pitch extraction algorithms, we could not determine a way to obtain a simple one-period based frequency measure. This measure is directly available in both CSpeech and BLISS.

With regard to spectral analysis, the main advantage that CSRE has over both BLISS and CSpeech is the capacity to generate a (rather spectacular) colour spectrographic display. A simultaneous power spectrum plot at any point in the spectrogram is easily obtained. There are a number of alternative approaches to spectrographic analysis and display available including analyses based on FFT (Fast Fourier Transform), Auto Regression and a Modified Covariance Method. One can listen to any portion of the spectrographic display using audio playback. By scrolling a cursor through the spectrogram, peak frequency values are displayed. One potential advantage is the high quality low cost colour hardcopies that can be generated with a Hewlett Packard Printjet. Finally, the speech synthesis capacity as well as the ability to work in one display environment while another is still active make CSRE an attractive speech analysis package.

Discussion

Overall we rate all three software packages very highly for their user-friendliness and ability to perform a number of relevant speech research measures. We think that all three might profit from more dialogue between developers, but we are extremely encouraged to find such powerful software packages available for microcomputers. We feel that it would be extremely useful if conversion utilities were available for making digital files from each of these systems compatible with the others. (As noted above, BLISS and CSRE are expected to be fully compatible in the near future.) These three packages are, in our opinion, the three main AT-based contenders, and a conversion utility would allow the user to exploit the particular advantages of each, without being stuck with a file-format that is unreadable by the other two.

To summarize briefly, the main advantages of each program are as follows. CSpeech is probably the most user-friendly with its transparent menu-driven commands and the ability to display multiple windows on a single screen. BLISS is most advantageous to those individuals who want maxi-

mum flexibility in defining analysis parameters. For individualized applications, BLISS is extremely powerful. Moreover, the audio experiment control program is quite useful. CSRE's distinct advantage is the colour spectrographic display, useful for both research and teaching needs. We hope that this review will be useful for those requiring information regarding speech analysis software for the AT-compatible machine.

CSpeech (US $1800)
Paul Milenkovic
Dept. of Electrical and Computer Engineering
University of Wisconsin
1415 Johnson Drive
Madison, WI 53706 U.S.A.

BLISS (US $250)
John Mertus
P.O. Box 3296
Wayland Square Station
Providence, RI 02906 U.S.A.

CSRE (US $350)
Donald Jamieson
Speech Communication Laboratory
Dept. of Communicative Disorders
Elborn College
Univ. of Western Ontario
London, Ontario N6G 1H1 CANADA

Address all correspondence to:
John Ryalls, Ph.D.
Ecole d'orthophonie et d'audiologie
Université de Montréal
C.P 6128, succ. A
Montréal (Québec) H3C 3J7 CANADA

References

Jamieson, D. 1989. Announcement: Canadian Speech Research Environment. London, Ontario: University of Western Ontario.

Jamieson, D., Nearey, T., & Ramji, K. 1989. CSRE: A speech research environment (Technical Note) *Canadian Acoustics / Acoustique Canadienne 17(4)*: 23- 35.

Mertus, J. 1989. *BLISS user manual.* Providence, R.I.: Brown University.

Milenkovic, P. 1989. *CSpeech: Computer speech waveform acquisition editing and analysis.* Madison, WI.: University of Wisconsin.

Read, C., Budner, E., & Kent, R. 1990. Speech analysis systems: A survey. *Journal of Speech and Hearing Research 33(2)*: 363- 374.