A MULTISENSOR DATA ACQUISITION AND PROCESSING SYSTEM FOR SPEECH PRODUCTION INVESTIGATION

Bernard Teston, Alain Ghio and Benoît Galindo

Laboratoire Parole et Langage-Université de Provence-CNRS

Aix en Provence France

ABSTRACT

The articulatory phonology study requires the simultaneous recording of the speech wave and as many articulatory parameters as possible. To this end, for many years,we have developed the integrated PHONART workstation for speech production analysis with the technologies available at the time. It consisted of a PC-compatible computer-driven workstation designed to record and process acoustic speech signals in relation to the corresponding physiological signals. These signals are obtained from various sources (flow-rate transducers, pressure transducers, position and movement gauges, electrodes, microphones, laryngophones, etc.). The parameters recorded with the PHONART system can be edited and processed with a new version of the PHONEDIT software. It is a signal editor that permits to edit, labelize and analyze various types of physiological signals in relation with the speech signal.

1.INTRODUCTION

The systematic investigation into the activity of different motor subsystems of speech and of articulatory acoustic correlations requires the simultaneous recording of the speech wave and as many articulatory parameters as possible. To this end, a six channel data acquisition prototype was first developed as part of a collaborative project between the speech research laboratory at Reading University and the IBM (UK) Scientific Centre, Winchester (Trudgeon et al, 1988). This system proved very useful for phonetic research as well as for the investigation of speech disorders (Hardcastle et al, 1988). Benefiting from this experience, we later developed at Aix an integrated workstation for speech production analysis (Teston et Galindo, 1990) enabling the recording, measurement, marking, segmentation and processing of sixteen physiological parameters in perfect synchronisation with the acoustic signal. Given its utility, we called it PHYSIOLOGIA. Since this time we have continuously improved it to begin the PHONART system which is a tool for researching physiological mechanisms of speech production. It consists of a PC-compatible computer-driven workstation, an acquisition system equipped with various transducers (flow-rate transducers, pressure transducers, position and movement gauges, electrodes, microphones, and laryngophones, etc.) and the signal editing and processing software PHONEDIT designed to record, edit, and process acoustic speech signals in relation to the corresponding physiological signals.

2. DATA AND CAPTURING SPECIFICATIONS

The movements of articulatory organs can be studied on three different levels. Analysis can deal with the neuromotor control of muscles involved in speech, their actual movements and the phenomena induced. The first level is essentially represented by electromyography; the second consists of direct movement analysis using videocinematographic imagery techniques or displacement transducers; the third one includes aerodynamic and acoustic phenomena, which evolve (as we move) through the vocal tract, and which, in their final state, produce the information- conveying speech signal.

Our objective is to analyze the greatest possible number of parameters which influence the production of speech. The difficulty of this task lies in the ability to simultaneously acquire data from different sources. Such acquisition depends on the quantity of information needed for the adequate description of the various parameters.

For theoretical reasons and practical considerations, a l msec time resolution to study articulatory organs movements in the vocal tract is sufficient. Signals from kinesiographic transducers fluctuate at the same rate as the movements of the articulatory organs, so a bandwidth of 2 kHz is quite sufficient in this case. Aerodynamic parameters require approximately the same bandwidth as kinesiographic signals. For electromyographic signals, a bandwidth of 2 kHz is also known to be sufficient.

Electropalatography (EPG) is a special case in this respect. Since multiplexing palatal contacts must be done within a short observation time interval, a 1 msec time resolution of the EPG frames is also quite sufficient. The synchronization link with video image acquisition systems is done by means of a 50 or 60 frames per second synchronization signal (European video standard and CCIR). The 2 kHz bandwidth is also sufficient here.

A bandwidth of 10 kHz is available for speech, in compliance with European ESPRIT SAM, project recommendations (Fourcin & al, 1989).

While a 12 bit A.D. converter resolution is sufficient for the physiological parameters dynamic it is not the case for speech signal, for which a 16 bit resolution is preferable mainly for a better dynamics, a good homogeny with SAM project and the « digital audio » sound.

Given all this, we make conclude that speech signal bandwidth must be four times larger than that of « physiological » signals.

3. DATA ACQUISITION SYSTEM

3.1. Acquisition interface

The acquisition interface comprizes four modules. The basic module features two acoustical inputs and two acoustical outputs, the aerodynamics parameters and one physiological input. One of the acoustical input can be replaced with four physiological inputs.

A second module contains the electropalatograph and eight physiological inputs The third module is a four channel electromyographic unit, and the last module is a new magnetoarticulometer (in evolution).
The interface is used with a 16 bit acquisition board Data Translation DT3016 featuring 32 inputs and a 200 kHz throughput. The maximal bandwidth of each input is hence of 3 kHz which is sufficient for physiological signals. Puting together four inputs we obtain a 12 kHz bandwidth which is also sufficient for acoustical signals. The basic module uses 16 inputs, the EPG 8 and the last 8 inputs are used for EMG and kinesiography.
For practical reasons of simplicity and thanks to the performances of modern PCs we use the interface with a single 200 kHz sampling frequency, regardless of the number of input parameters.

3.2. Acoustical inputs

The speech signal can be recorded via two acoustical inputs:
-from an electrodynamic, symmetric or asymmetric microphone, an electret microphone (with a phantom power supply furnished by the system). It is possible to calibrate it and other types of microphone to make sonometric measurements.
-from a line input for signals recorded on DAT tape recorders, or generated by an external transducer such as the electrolaryngograph.
The acoustical inputs are equipped with an anti aliasing filter (8 orders Butterworth) with a 12 kHz cutoff frequency. If the number of physiological inputs is not sufficient, the second acoustical input can be replaced with four physiological inputs.

3.3. Physiological Inputs

The eight physiological inputs have a frequency bandwidth equal to one fourth of the acoustical input bandwidth. Two of them can be equipped with anti-aliasing filters identical to the one used for the acoustical inputs, with a fixed cutoff frequency of 3 kHz. Their maximum level is plus or minus 10 volts.

3.4. Electropalatograph (EPG)

The EPG system consists of a special coupler device in the acquisition interface which works with the "Reading" EPG palatal plates marketed by the Millgrants Company (UK). It can be used to visualize the points where the tongue touches the palate. Tongue contact is detected by 62 electrodes placed in various positions on a palatal plate. The coupler device transforms EPG digital data in eight 8-bits analog signals which use one physiological input.This data is thus synchronous with data corresponding to acoustical and physiological parameters. The synchronisation error with the acoustical signals is less than 1 ms.

3.5. Remote control system

The remote control system can be used to acquire physiological and acoustical parameters on the computer by:
- The starting and stopping of acquisition in mode AB or C (international instrumentation standard) upon receipt of signals from manual action, from an programmable acoustic signal threshold level, or from a device outside the interface.
- Remote control of tape recorders (DAT) or devices outside the interface.

It is possible to synchronise video recordings with the acoustical and physiological signals through an extension of the remote control system.

3.6. Acoustical output

The interface has two acoustical outputs for listening to speech signal files, equipped with an amplifier for use with a loudspeaker or earphones. The output signal reconstructing filter is identical to the one used for the acoustical inputs and has the same frequency cutoffs.

4. ACQUISITION DRIVER

This set of programs handles the acquisition of the speech signal and of the various articulatory parameters.The acquisition program includes a data acquisition module in which the user defines the instrumental configuration. The experimenter selects a channel for each parameter. It indicates to the system the input numbers and types of signals used (acoustical or physiological, oral and nasal airflow, oral phonogram, etc..). The acquisition configuration can be saved in a dedicated file. The level for each signal is adjusted and the user is informed by bargraphs if any saturation occurs. A special window on the screen allows to adjust the threshold of the EPG. Automatic calibration is performed for aerodynamic and kynesiographic signals. The sampling frequency is fixed to 200 kHz regardless of the number of input parameters.
The acquisition file headers are in RIFF format (an ISO standard registered under the name of EAIFF85). The file headers contains several fields, some of which are optional. The fields include: acquisition specifications (optional): number of channels, number of samples, sampling frequency, resolution (8, 12 or 16-bit), maximum value, minimum value, zero. Signal specifications (mandatory): signal code, signal name, number of samples, sampling frequency, largest value, smallest value, maximum calibration, zero calibration, unit of measurement. The signal (mandatory). Additional information (optional): creation date, comments on corpus, key words, signal source, acquisition software, copyright, recording and storage place, user's name, etc. Subject date (optional): name, age, sex, native language, etc...
Acquisitions are generally started and stopped manually by the experimenter or the speaker. The acquisition duration can be very long. It is only limited by the size of memory. The raw acquisition data file containing interspersed segments of data pertaining to the various parameters is split down into as many files as there are physiological signals. The acoustic signals require an additional mixing operation. The parameter files are then saved on disk.
In practice, acquisition is done one sentence at a time.

5. MEASUREMENT DEVICES

The PHONART environment offers a set of transducers and measurement devices which are particularly suited to the study of the mechanisms of speech production. They are normally used on the EVA medical work station for speech and voice pathologies (Teston et Galindo 1995).

5.1. Aerodynamic transducers

The aerodynamic transducers perform what we call the aerophonometric function, i.e. they measure the air pressure levels in the vocal tract and the resulting inhaled and exhaled,

oral and nasal airflow rates, as a function of the movement of articulators.
Oral airflow transducer:
The oral airflow rate is measured by a grid pneumotachograph (PTG) with a low dead volume, good linearity, and a wide range on 2 scales: 2 and 10 dm3 per second (or liters/second). The interface between the subject's face and the transducer is achieved via flexible silicone rubber mouthpiece (Teston 1993).
Nasal airflow transducer:
The nasal airflow transducer is identical to the buccal airflow transducer. Its location and shape were designed to guarantee maximum nasal air evacuation and measurement accuracy. The nasal airflow rate is measured in the nostrils using silicone nosepieces of various sizes. The measurement scales and ranges are the same as for the oral airflow transducer.
Intra oral and subglottic pressure:
This parameter are measured with pressure transducers on 2 scales: 20 and 100 hPa (hectoPascal).

5.2. Instantaneous pitch meter
The pitch meter measures the instantaneous vibration frequency of the larynx with great acuracy, period by period, based on the speech signal acquired by microphone or by the laryngeal transducer. It operates in real-time and has four measurement scales: 250, 500, 1000, and 2000 Hz.

5.3. Intensity detector (sonometer)
The intensity detector is a sound level meter. It measures the logarithm of the RMS of the speech signal. Its integration time constant is 10 ms (50 ms can be selected for very low, male voices).
Normalized frequency weighting "A" is also available.). It can be calibrated for two microphones furnished with the system.

5.4. Respiration transducer
This device is a piezoresistive « pneumotrace » which measures the difference in total abdominal circumference. So, it permits to record the respiratory rythms during speech production. It is connected to the physiological input of the basic module.

6. SIGNAL EDITION AND DATA PROCESSING
The signal editing module is the PHONEDIT software system, which runs in the Microsoft WINDOWS 98 environment. It is used to visualize, segment, mark off, measure, and process the recorded parameters. PHONEDIT also recognizes most current types of files: ACCOR Edit System (Reading), SAM (GERSON, BDSONS, etc.), ILS, Microsoft WAWE (Multimedia on PC); KAY CSL, SIGNALYZE, SOUND WAWE, ACSII Format, RAW and RAW UNIX. It is possible with a special program (DLL) to enable the design of custom files.
Icons, pop-down menus, and numerous utility programs make the program very easy to use.
Since PHONEDIT runs under WINDOWS, a spreadsheet such asEXCEL or a data base management system like ACCESS can be used directly to process the data output by the PHONART system.
For labelling and data reduction operations, the PHONEDIT software can be used without the acquisition interface. It is possible in this case to listen to the PHONART files via WINDOWS multimedia boards like SOUNDBLASTER.

6.1. Measurement operations
One-cursor operations:
By placing a single cursor on a curve, the user can measure linear or logarithmic amplitudes, calculate the spectrum (FFT or otherwise), visualize tongue-palate contacts (EPG), or organ movements (movetrack) or insert a label or segmentation marker (alphanumeric, symbolic, or phonetic alphabet) on a line reserved for that purpose
Parameter amplitudes can be calibrated by the calibration signal (zero and maximum) generated by certain measuring devices, by specifying the parameter maximums and units to the computer, or automatically, if our custom-made instruments equipped with a calibration system are used.
Two-cursors operations:
Between two cursors, the user can zoom and scroll measure durations, count events. listen to the signal and calculate the linear or logarithmic difference, the integral, mean, standard deviation, or variation coefficient. It is possible to create a cursor with a variable length (zone cursor) wich defines a moving window. In this window different statistical treatments can be made as between two cursors.

6.2. External functions
External functions are operations applied to parameter curves. They consist of a dynamic link program library (DLL) performed by three modules.
Computation module:
This module executes the following operations: sum or difference of two curves, absolute value, integration (variable-length steps), powers, roots, logarithms, RMS, quadratic spline, level quantification, variable time shifts, phase inversion, synchronous mean, EPG sum elevation and translation, and EMA (Movetrack or Articulograph) movements interpretation. A statistical library can be used with shimmer and jitter in adition to the treatments of the zone cursors.
A detailed source file such as DLL squeleton is avalaible to allow users to develop their own specific operations.
Acoustic analysis module:
This module performs frequency-amplitude analysis (FFT, LPC. 1/3 octave, critical bands, long-term spectrum) and time-frequency-amplitude analysis (wide or narrow band sonograms followed by formants).
Pitch analysis module:
It is composed of two methods of pitch detection, based on the COMB and AMDF.
The melodic curves can be modelled by spline and automatic detection of target points.
All measurement results (time, amplitude, frequency, statistic etc..) can be saved to a table processed by commercial spreadsheet programs (EXCEL,etc..). A utility program allows the printing of the different screen windows, which can be copied, by "drag and drop" in various editor progams such as WORD.
Screens can be printed in colour or black and white on any printer supported by WINDOWS.

7. CONCLUDING REMARKS

Our ongoing concern while designing and developing this workstation was to offer the user as much operating ease as possible. This was achieved by combining windows, and pop-down menus accompanied by many utility programs giving the user in charge of data processing a kind of workability that is unknown elsewhere. Never before has it been possible to access such a complete set of data on articulatory processes. For the researcher, knowing at any given instant, for example, the magnitude of oral and nasal airflow, intra-oral pressure, lip opening and rounding, the location of the tongue on the palate, and the spectrum of the acoustic signal is a highly valuable capability (Nicolaidis et al, 1993)

We hope the use of the PHONART workstation will contribute effectively to improving our knowledge of articulatory processes.

REFERENCES

Fourcin, A.J., Harland, G., Barry, W., and V., Hazan, (eds)(1989), *Speech Input-Output Assessment: multilingual methods and standards*. Chichester: Ellis Horwood.

Hardcastle, W., Jones W., Knight, C., Trudgeon, A. and G., Calder (1989), "New developments in electropalatography: a state of the art report". *Clinical Linguistics & Phonetics*, vol. 3, 1-38.

Hardcastle, W.J., Morgan Barry, R.A., Clark, C.J., and M., Nunn, (1988), "The use of instrumental phonetics in assessment and remediation". In stengelhofen, J. (et.), *Cleft Palate: case studies in the treatment of communication disorders*, Edinburgh Churchill Livingstone, pp. 136-164.

Nicolaidis, K, Hardcastle, W.J., Marchal, A and N. Nguyen-Trong (1993), "Comparing phonetic, articulatory, acoustic and aerodynamic signal representations". in M. Cooke, S. Beet and M. Crawford (Eds). *Visual Representations of Speech Signals*, Chiches ter: Wiley & Sons, pp. 55-82.

Teston, B. and B. Galindo (1990) "Design and development of a work station for speech Production Analysis". *Proc. VERBA 90. International Conference on Speech Technologies*. Rome, pp. 400-408 .

Teston, B., (1993), " A new air flowmeter design for the investigation of speech production". *Proc. Eurospeech 3*, Eur. Speech Com. Ass., Berlin, sept. 93, 405 409.

Teston, B. et Galindo, B., (1995), "A diagnostic and rehabilitation aid workstation for speech and voice pathologies", *Eurospeech 4*, Eur. Speech Com. Ass., Madrid, sept. 95, 1883-1886.

Trudgeon, A., Knight. C., Hardcastle, W.J., Calder G. and F. Gibbon (1988), "A multi-channel physiologicial data acquisition system based on an IBM PC, and its application to speech therapy". *Proc. Speech'88 (7th FASE Symposium)*, pp. 1093-1100.