Proceedings of the Institute of Acoustics

THE VOICE-STICK: A VOICE CONTROLLED SEGA JOYSTICK

James A S Angus and Robert Smart

Department of Electronics, University of York, York, England YO1 5DD

1. INTRODUCTION.

Children, and adults, who have speech or hearing disabilities are often given exercises by speech therapists that involve making a lot of, frankly boring, repetitive noises. The patient may also feel a bit silly while uttering these noises to a blank wall. How much better it would be if one could combine both therapy and pleasure by using these noises to control an exciting and interactive video game. One could do one's speech exercises, and relieve ones stress, by shouting at Sonic The Hedgehog or Super Mario! The purpose of this project was to design a voice controlled device that could replace the joystick of a standard, and therefore readily available, game machine. This would allow both a cost effective implementation, and access to a wide variety of games to practice one's voice on.

The paper will describe the results of this project. It will focus particularly on the problems associated with interfacing speech parameter extraction circuitry with the interface protocols of the game machine.

2. THE PROJECT.

The aim of the Voicestick project was to design a games console joystick that can be controlled by different vocal sounds. The project can be split into two closely linked goals. The first of these was to produce a unit that will enable games console games to be partially or completely controlled by different vocal commands; a voice controlled joystick. The second part of the project was to tailor the design so that the vocal sounds it responded to were useful for the vocal exercises used in speech therapy. The main purpose behind this second part of the project was to assist the vocal exercises often used in speech therapy.

Speech Therapy is a specialist area of work and because of this the equipment designed to assist speech therapists in their work tends to be expensive. This is mainly due to large development and production costs coupled with low production quantities. The Voicestick project aimed to satisfy this need by producing an item that is relatively cheap and can be easily produced in large quantities. This need for large quantities can be generated by marketing the Voicestick as a novelty item to be purchased by people already owning the console system, as well as a speech therapy aid. The large production quantities would in turn reduce the cost of the product.

3. DESIGN CONSIDERATIONS.

It was decided that the prototype voicestick was to be made compatible with the Sega Megadrive games console. The Megadrive system was chosen because of the system's popularity as it is currently the best selling games console on the market. The popularity of the console will provide a large number of potential customers. However on completion it is

possible for the market of the Voicestick to be increased by producing versions that would be compatible with other joystick-driven devices, such as other games console systems, and computers.

Speech therapists use a number of different exercises to assist them in their work. These assist with a variety of speech problems. The more common exercises involve repeating vowels, alternating between voiced, unvoiced, and mixed sounds, and varying the points of articulation in the vocal tract. The latter of these tends to be an exercise that is tailored to the individual and therefore more difficult to implement as a hardware project.

The vocal exercise set that has been chosen for the Voicestick is changing between voiced, unvoiced and mixed voicing sounds. One of the objectives of the chosen exercise would be to practice the control and timing of opening and closing the larynx. The Voicestick would respond to pure voiced sounds, pure unvoiced sounds, and sounds that consist of a mixture of both voiced and unvoiced energy.

The Voicestick will be used primarily by children and teenagers and therefore it is necessary that the unit should be able to withstand any mistreatment that it may receive. Also because of the possible use by young children it is necessary for the unit to be simple to use.

Thus the ideal specifications for the Voicestick are as follows:

- Be fully or partially controlled by vocal sounds.
- Be effective and useful in the application of speech therapy.
- Be compatible with the Sega Megadrive.
- Be cheap to produce.
- Be easy to produce in large quantities.
- Be easy to use.
- Be suitable for use by children.

4. DETERMINATION OF BASIC METHOD.

Several basic means of determining whether a vocal sound is voiced or unvoiced have been developed, including techniques using Neural networks, Pattern recognition [1], Auto-correlation [4], Various statistical methods, and other methods under the general categories of digital and analogue signal processing. Most of the methods are unsuitable for this particular application for various reasons. Many are unsuitable simply because of the cost of implementation. An obvious example is that some of the methods require the use of microprocessors. Other methods are unsuitable because of the time taken to make a decision. For this application the response time of the system needs to be of the same order as the human response time (in the order of milliseconds). The method of direct frequency separation by analogue filtering, (the one chosen) has the advantages of being quick to respond, simple and cheap to implement but still adequately reliable in its action.

4.1 Preliminary Voice Testing.

To find the frequencies at which the analogue separation filters should be set at, voice samples from six people of varying age and sex were taken. The samples contained various voiced and unvoiced sounds which where then analysed with the help of the Apple Macintosh software "Analyser". The spectral energies of the samples were measured to show relatively where the energy of each sample was concentrated.

The results of these tests on the sampled voices showed that voiced sounds have frequency components between 100Hz and 2kHz and unvoiced sounds have frequency components between 3kHz and 16kHz. However the spectral energy information shows that the majority of voiced speech energy occurs between 300Hz and 1kHz, and the majority of unvoiced energy occurs between 5kHz and 14kHz. Information in Knorr [3] confirms these figures.

The 'Analyser' analogue filter simulation was used to find suitable filter responses for each of the separation filters. The corner frequencies for the voiced low pass filter, and the unvoiced high pass filter where chosen at 1kHz and 5kHz respectively due to the spectral energy results. It was found using the filter simulation that suitable attenuation in the stop band to allow reliable decision can be achieved if, for the low pass (voiced separation) filter, a sixth order Butterworth was used. For the high pass (unvoiced separation) filter an eighth order Butterworth is sufficient. Knorr [3] suggests that a high pass sixth order Chebychev filter is adequate for unvoiced separation. This was tested using the filter simulation and found to be true. Because of its smaller circuit realisation the Chebychev solution was chosen for the final implementation.

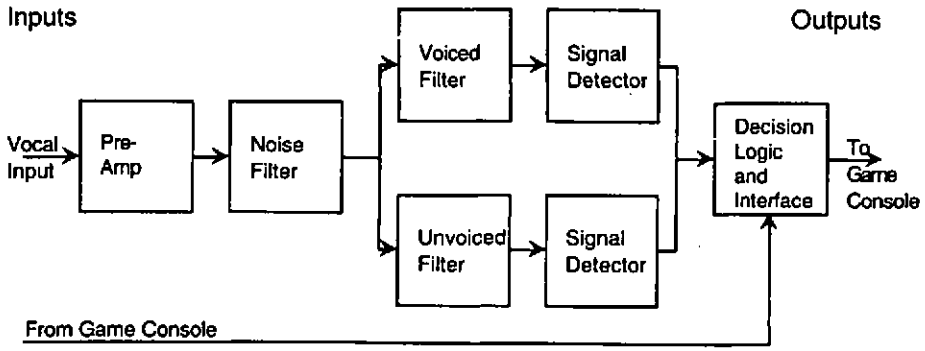

Figure 1 Block Diagram of the Voicestick

5. TOP LEVEL DESIGN

The design consists of four basic parts, the microphone and pre-amplifier, the voice parameter separation filters, the decision circuits, and the Sega interface circuit and is shown in figure 1.

The basic principle of the design is that the unprocessed vocal sound enters via the microphone, undergoes some pre-amplification and gets passed into the parameter separation section of the circuit. By filtering, the vocal sound is split into the two channels required for voiced, unvoiced, mixed decision. The two channels feed decision circuits that indicate whether the amplitude on the channel is above or below a pre-set threshold. The digital output from the decision circuits is finally passed though a circuit to interface with the Sega.

In addition to the above an extra filter has been incorporated to reduce of the effect of background noise and increase the reliability of operation.

5.1 The Noise Reduction Filter.
The noise reduction filter is designed to reduce noise that occurs outside the speech frequency band (300Hz - 14kHz). The filter is in two parts, one part removes low frequency noise and the other the high frequency noise.

The low frequency filter is concerned with reducing the noise with frequencies less than 300Hz. Background noise at these frequencies has a relatively low amplitude and because of this all that is required is a second order filter.

The upper frequency limit of the useful speech frequency band is 14kHz, and the upper limit of the microphone frequency response is specified at 8kHz. It was found that, even with the band pass nature of the microphone, the reliability of the voiced, unvoiced detector is increased if a High frequency noise reduction filter is incorporated. The filter that has been chosen is the similar to that used by Knorr [3], a fourth order low pass filter with a Butterworth response. The corner frequency of this filter is set at 12kHz rather than 14kHz so that television line-frequencies are removed as well as other high frequency noise.

5.2 Voice Parameter Separation.
The output signal from the noise filters is fed into the parameter separation module of the Voicestick. The parameter separation module uses two filters to split the speech frequency band into two smaller bands, one for voiced (300Hz-1kHz) and the other for unvoiced (5kHz-12kHz), as discussed earlier. Both of the filters are realised using Sallen Key second order sections because of their ease of implementation.

5.3 The Switching Circuit.
The switching circuit takes the varying signal on the two channels and gives a digital output suitable to drive the CMOS logic gates in the interface circuit. There are three stages to the switching circuit, these are full wave rectification, integration (or low pass filtering) and a threshold comparator. The full wave rectification is performed by a standard precision rectifier. The output for the rectifier is then smoothed by an integrator. This is done by a simple passive low pass filter with a comparatively short time constant. The final section of this stage is the threshold comparator circuit. For this a standard operational amplifier in open loop mode is used, as the required speed of switching is comparatively slow.

6. INTERFACING WITH AND CONTROL OF THE CONSOLE.

Two methods of providing all the inputs to the games console were considered:

1 Replace the standard games pad and generate all the required inputs within the Voicestick.

2. Design the Voicestick as a plug-in between the standard games pad and the console.

The second of these options was been chosen, because of the cost difference. The cost of incorporating all the extra buttons and circuitry to provide all the outputs would have significantly increased the cost of the product.

The interface circuit is required to take inputs from the Sega game pad and the voiced/unvoiced detector and provide outputs to drive the games console. The interface can be split into two sections. The first takes the outputs form the voiced unvoiced detector and uses logic to generate the three required functions. The second section of the interface circuit formats the

outputs into the same form as the games pad. In order to perform both these functions we need to understand how the Sega game pad interface works.

6.1 The Sega Game Pad

The Sega Game Pad contains the following circuit (figure 2), the 74HC157 chip is a CMOS multiplexer chip.

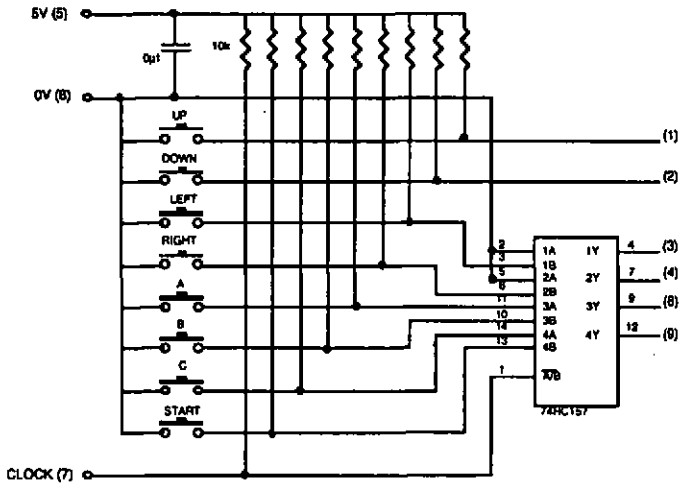

Figure 2 The Sega Game Pad

The interface is very simple, consisting of two direct input switches (**up** and **down**) and four inputs that represent multiplexed switches (**left**, **right**, **A**, **B**, **C** and **Start**). The multiplexed switches are selected by a clock signal from the game console that is strobed every 20ms, implying that the fastest reaction time available from the system is 20ms. This would be consistent with the fact that the television picture, on which any reaction might take place, can only be updated at this rate.

The three functions that are generated from the output of the voiced unvoiced detector are **left**, **right** and the function button **B**, which is used for the jump command. The second stage of the interface circuit is based on the circuit in the game pad. There are two types of outputs from the game pad some are the direct output from a grounded switch and the others are the outputs from a multiplexer circuit. The Voicestick interface circuit takes the all the outputs from the games pad, replaces the functions **left**, **right** and button **B** and passes the new outputs into the console. To do this the interface circuit uses a multiplexer configured in the same way as the one in the game pad. The multiplexer in the interface circuit is clocked by the same clock as the game pad. A second multiplexer is present in the interface circuit to provide the Voicestick's "transparent mode", where the game pad can be used in the normal way. Figure 3 shows the circuit of the console interface.

THE VOICE-STICK: A VOICE CONTROLLED SEGA JOYSTICK

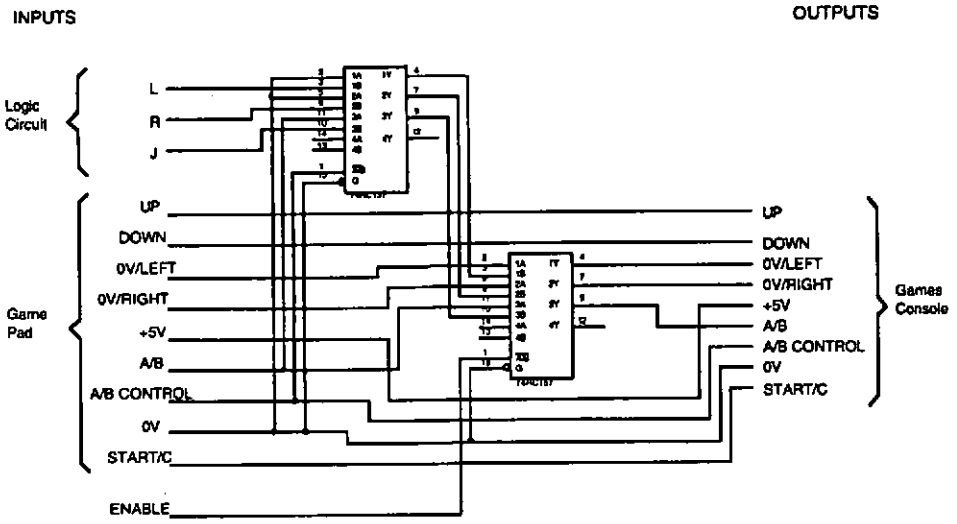

Figure 3 The Sega Interface Circuit.

7. THE EXTERIOR DESIGN.

The exterior of the Voicestick has been designed to optimise the performance of the microphone. This has been done by the use of two surfaces to increase the sensitivity, frequency response and directivity of the microphone. The surfaces used are the top of the box and a second surface of the same size attached to the top by a hinge.

The final consideration of the surfaces used is the physical positioning of the microphone to get the flattest frequency response. The flattest frequency response can be obtained by placing the microphone one third of the way off the centre of the junction between the two surfaces. To summarise, the use of two surfaces of the size of the top of the Voicestick box has the following effects, from Bartlett [2]:

- An increase in the acoustic sensitivity of the microphone by 12dB at all frequencies above 880Hz.
- An increase in the directionality of the microphone at all frequencies above 320Hz.
- A additional minor peak in the microphone frequency response at 1.68kHz.

8. DESIGN IMPROVEMENTS.

One design improvement would be to improve the function of the voice controlled **B** button ("Jump"). The problem arises from the fact that, when using the standard game pad, the length of time that the button is pressed affects the function of the button. For example in "Sonic The Hedgehog" if the button is tapped then the character performs a small jump, whereas if the

button is pressed for a longer time a higher jump is performed. When the voice control is used it is often difficult to "jump" high enough, this is because most of the time the output of the comparator switches for only a short time, which is the equivalent of tapping the control pad button.

To rectify this problem there are a couple of possible solutions. The first is to simply incorporate a monostable multivibrator between the output of the logic circuit and the interface multiplexer. The monostable would lengthen the short pulse outputs from the comparator into a longer pulse which would be the equivalent of pressing the button on the game pad for longer. The second and preferred solution is to add a schmitt trigger to generate the B input to the interface circuit. The use of a schmitt trigger to generate the B function will cause the output pulses to be longer because of the switching characteristics of the schmitt trigger. The length of the output pulse can be altered by changing the schmitt triggers hysteresis band.

There are several other areas where the Voicestick can be extended to provide more choice of vocal exercise for the speech therapist to use. However these extensions would increase the price of the unit and change the target market towards the specialist speech therapy market. This would in turn significantly increase the cost, because of the lower production quantities.

9. CONCLUSIONS.

The principal aim of the project, that is to produce a unit that is "fully or partially voice controlled by vocal commands", has been met, and a partially voice controlled unit has been designed and produced. The voice controlled unit has also been successfully interfaced to the Sega Megadrive games console.

The unit is easy to use as it has a few simple controls. This, and the fact that the unit is fairly robust and resistant to the general abuse that it would be expected to suffer, makes the unit suitable for use by children. Finally the unit is suitable for production in large quantities because there are relatively few outside connections, and the circuitry fits simply onto printed circuit board. If the demand was great enough the circuit could easily be transferred to surface mount printed circuit board as all the components used are available in surface mount form.

The Voicestick adds a new dimension to video games and speech therapy in that if you can hum it, you can play it.

10. REFERENCES.

[1] B. S. Atal, A pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Application to Speech Recognition," IEEE Trans. Acoust., Speech, Signal processing, Vol. ASSP-24 no. 3, June 1976.

[2] Bruce Bartlett, "The PZM Boundary Booklet - A Basic Primer and Experimenter's Guide," Appendix C Microphone Manual, Design and Application by David M. Huber. Pub. Howard W. Sams and Co.

[3] Siegfried G. Knorr, "Reliable Voiced / Unvoiced Decision," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-27 no. 3 June 1979.

[4] D. A. Krubsack and R. J. Niederjohn, "An Autocorrelation Pitch Detector and Voicing Decision With Considence Measures Deveolped for Nose-Corrupted Speech," IEEE Trans. on Signal Processing, Vol. 39 no. 2 February 1991.