The Chatterbox

A simple speech synthesizer for demonstration and amusement

by Ian H. Witten, M.A., M.Sc., Ph.D., M.I.E.E. **and Peter H. C. Madams,** B.Sc., M.Sc.

Department of Electrical Engineering Science, University of Essex.

The device described is a hand-controlled, electronic model of the acoustic properties of the vocal tract, and was built to illustrate the physiological and acoustic nature of speech. Although designed as a portable demonstration and lecturing tool it makes a fascinating toy for adults and children alike, and has been used as a stimulus for retarded and autistic children. After discussing the nature of speech and the mechanism of electronic speech synthesis, the authors explain the design principles of the Chatterbox and in a later article will give further circuit details and instructions on how to make it talk.

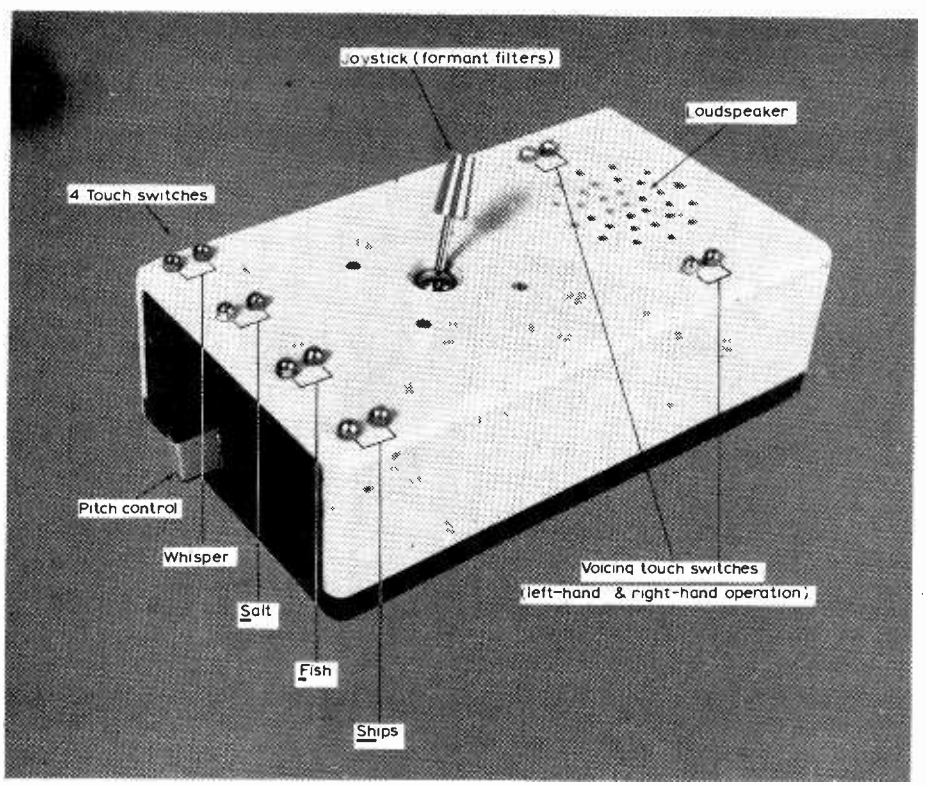
PEOPLE SPEAK by using their vocal chords as a sound source, and making rapid gestures of the articulatory organs (lips, tongue, mouth, etc.). The resulting changes in shape of the vocal tract allow the production of the different sounds that we know as the vowels and consonants of ordinary language. For several years it has been possible to simulate the action of the vocal tract electrically, using a device similar to an electronic organ to produce sounds with the same character as those of human speech. The first of these "speech synthesizers," built in the early 1950s, comprised many racks of equipment, consumed a lot of power, and cost a great deal of money. Now, however, with the advent of cheap integrated circuits, it has become possible to build simple, compact, and quite inexpensive synthesizers, without sacrificing the ability to produce the full range of speech sounds.

Of course, to make the ever-changing patterns of speech, a synthesizer needs some form of continuously varying control, and just as there are many vocal tract organs involved simultaneously in speaking, so it is necessary to control several parameters of the synthesizer at once. Most speech research laboratories nowadays use a digital computer to manipulate the control signals for their synthesizers. However, for the purposes of informal experiments with speech or just to learn about the sounds we make, a pair of hands will suffice – with the added advantage that the operator can use his long-standing experience with real speech to mould the sounds into voice-like ones.

By way of illustration of these points, we have built a small, manually controlled speech synthesizer, suitable for home construction – the "Chatterbox" (Fig. 1). In experienced hands it can be encouraged to utter recognizable words and phrases ("hello," "how are you," etc.), while even a complete novice can make it generate a great variety of astonishingly different noises, all of which are immediately recognizable as speech-like. The Chatterbox was originally designed as a portable demonstration and lecturing tool for illustrating the different sounds of speech and how they can be synthesized, but we quickly found that the fascination of artificial speech makes it a successful and compelling toy for adults and children alike. As a hand-controlled, electronic model of the acoustic properties of the vocal tract, it provides a natural feel for the growing science of phonetics — the study of what people do when they are talking and when they are listening to speech.

This first article discusses the nature of speech and the mechanisms of electronic synthesis. The Chatterbox design is described later, with some circuit details, as a concrete example of the implementation of a speech synthesis system.

Fig. 1. *The Chatterbox, showing the controls for hand operation.*



The anatomy of speech

The so-called "voiced" sounds of speech – like the sound you make when you say "aaah" – are produced by passing air up from the lungs through the larynx or voicebox, which is situated just behind the Adam's apple. The vocal tract from the larynx to the lips acts as a resonant cavity, amplifying certain frequencies and attenuating others.

The waveform generated by the larynx, however, is not simply sinusoidal. (If it were, the effect of the vocal tract resonances would merely be to give a sine wave of the same frequency but amplified or attenuated according to how close it was to the nearest resonance.) The larynx contains two folds of skin – the vocal cords – which blow apart and flap together again in each cycle of the pitch period. The pitch of a male voice in speech varies from as