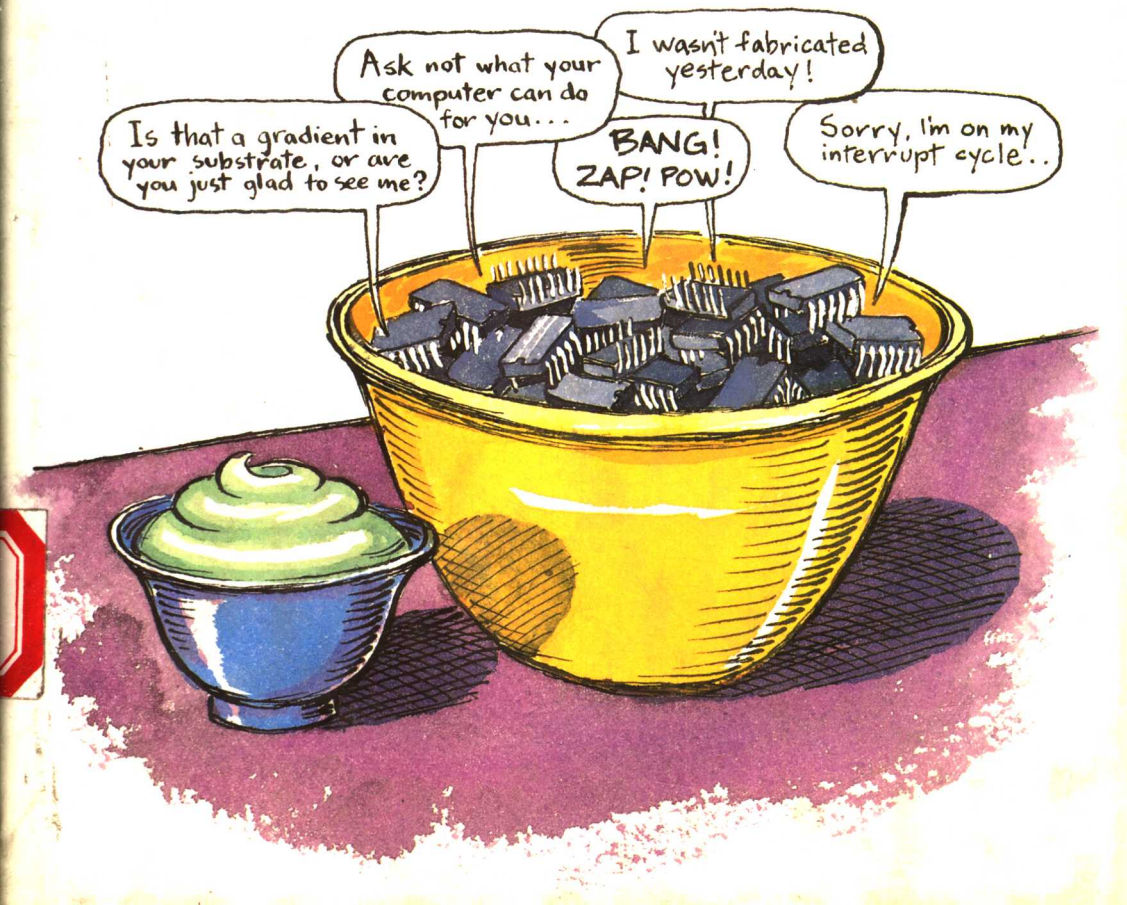
National's Semiconductor Technology Series

TALKING CHIPS

IC SPEECH SYNTHESIS

NELSON MORGAN

With Illustrations by Beatrice Benjamin



TALKING CHIPS

Nelson Morgan

***With special contributions from* Jake Buurma *and* Lloyd Rice**

***Cartoons by* Beatrice Benjamin**

McGraw-Hill Book Company

New York St. Louis San Francisco Auckland
Bogotá Hamburg Johannesburg London Madrid Mexico
Montreal New Delhi Panama Paris São Paulo
Singapore Sydney Tokyo Toronto

Library of Congress Cataloging in Publication Data

Morgan, Nelson.
Talking chips.

(National's semiconductor technology series)
Bibliography: p.
Includes index.
1. Speech synthesis. I. Buurma, Jake. II. Rice, Lloyd. III. Benjamin, Beatrice. IV. Title.
V. Series.
TK7882.S65M67 1984 621.3819 83-14914
ISBN 0-07-043107-8

1234567890 DOCDOC 89210987654

ISBN 0-07-043107-8

The editors for this book were Harry Helms and Susan Thomas, the designer was Jules Perlmutter, and the production supervisor was Thomas G. Kowalczyk. It was set in Primer by Santype-Byrd.

Printed and bound by R. R. Donnelley & Sons Company.

FOREWORD

The development of algorithms that can be used in computers to simulate human speech has been a research topic for a number of years. Until recently, these algorithms could only be implemented with the fastest special-purpose computers. This restricted their use to demonstrations performed in research laboratories. The advances of integrated circuit technology have now made it possible to implement the high-speed calculations on a single chip and have therefore dramatically opened up the number of potential applications for which speech can be used. In this book, Morgan has demystified the technology that makes it possible for integrated circuits to talk. I am sure he hopes, as I do, that a number of the readers of this book will take the next step of determining which speech applications will really be useful.

In the near future, the same technology that makes speech synthesis possible will be applied to speech recognition, which will give the computers the capability of listening. We can only hope that Morgan will unravel the complexities of these "listening chips" in the same humorous way as he has done it for the "talking chips."

R. W. Brodersen
Professor, U. C. Berkeley

PREFACE

Synthetic speech has been discussed at length in a number of excellent texts, some of which are listed at the end of Appendix C. It is only recently, however, that high-density integrated circuit (IC) techniques have brought speech technology to the consumer. The inquisitive user of speech products will want to understand the basics of these gadgets. This book is an attempt to bridge the gap between graduate-level texts and marketing hype.

Talking chips have developed from a fusion of disciplines including mathematical signal processing, microcomputers, electronics design, and acoustics. It is unlikely that many readers will have a strong background in more than one of these areas. Consequently, I have attempted to emphasize concepts and examples over theory and rigor.

The Introduction and Chapters 1 through 3 cover the topic of speech synthesis systems. The Introduction suggests some whys and wherefores for the technology. Chapter 1 provides a basis for the speech theory. Chapter 2 shows the range of hardware options to implement these theoretical notions. Chapter 3 is a collection of case histories of real speech chips.

Chapters 4 through 6 describe the process of parameter generation for speech synthesis systems. Chapter 4 discusses techniques for analysis of natural speech. Chapter 5 introduces phonetic rules for automatic synthesis from text or phonetic symbols. Chapter 6 covers the commonly ignored audio and acoustic considerations for recording and playback of speech.

Chapter 7 is a brief overview of the projected future for this technology.

Appendixes A and B are provided to explain a few background points about signal processing. Appendix C outlines the creative process of vocabulary generation for fixed-vocabulary synthetic speech.

Jargon has been deemphasized where possible, but there were certainly places where it had to be used. For example, Chapter 3 assumes some

familiarity with the more common IC and microprocessor terms. The intrepid reader without appropriate background can make use of the glossary at the back of this book.

I had a lot of fun writing this book. I hope you enjoy reading it.

Nelson Morgan

ACKNOWLEDGMENTS

There are several people who were major forces toward the completion of this book. First and foremost was Bryan Costales, who contributed countless hours reading and correcting my scribbles in an attempt to make them intelligible. Lloyd Rice and Jake Buurma contributed the raw material for Chapters 3 and 5. Artist-at-large Beatrice Benjamin provided the drawings that may have saved this from being just another technical book. Finally, Harry Helms must be congratulated for pushing me into writing the book.

Many others assisted as well. Excuse me if I miss some of you, but here goes. Thanks to:

Juniper
Max Hauser
Hy Murveit
Steve Pope
Joe Santos
David Isenberg
Tom Frederiksen
Bob Brodersen
Joe Costello
R. F. Morgan
D. Thrall of the Telephone Museum
B. J. Morgan
Geri, Ben, and Soma
My ex-wives

and to the myriad researchers who have shared their experience so freely in casual conversations.

CONTENTS

INTRODUCTION

FOR RELEASE
in Morning Papers
January 6, 1939

THE VODER

An electrical device which, under control of an operator at a keyboard, actually talks was demonstrated on January 5 at the Franklin Institute, Philadelphia. Known as the VODER, it is a development of Bell Telephone Laboratories as a scientific novelty to make an interesting educational exhibit for the Bell System's displays at the San Francisco Exposition and at the World's Fair in New York. It is built, except for its keys, entirely of apparatus used in everyday telephone service.

The VODER creates speech.

The VODER[1] (Voice Operation DEmonstratoR) was not a standalone talking machine. It was played like an instrument by a trained operator. The resulting acoustic output, however, bore a striking resemblance to the human voice. Someone would say, "How are you today, VODER?" The operator would manipulate keys and pedals and VODER would respond, "Fine. How are you?" This was entertaining and intellectually enticing as it showed how far the science of that time had come in its understanding of the production of speech.

The principal architect of this electrical marvel was a Bell Laboratories researcher named Homer Dudley. He did not, of course, create his VODER in a conceptual vacuum. Thinkers as far back as the Greeks have been intrigued with the idea of talking automata. Actual mechanical talking machines were built as long ago as 1779 by innovators like Kratzenstein and Von Kempelen. Vibrating reeds meant to simulate our vocal cords were connected to resonant chambers analogous to our vocal tract. While the speech quality must have been quite poor, these machines reputedly did work for discrete sounds.

It was not until Homer Dudley's VODER, however, that a machine was able to produce connected speech. Today, purely electronic and automatic devices, some of them contained on a single integrated circuit, can produce speech. Surprisingly, the fundamental theories involved in speech synthesis have not changed greatly from those used at the time of the VODER.

There has been an incredible upsurge in speech-synthesis design activity over the last few years. As of this writing, over one dozen speech-synthesis IC's have appeared. Virtually every electronics company of any size now employs a speech group of some sort. Why has this happened now? The high density and speed of modern *very large-scale integrated* (VLSI) circuits have given us the ability to build complex systems on a single chip of monolithic silicon. Manufacturers have been looking for something to do with all of that capability.[2]

Prior to this development, speech research was relatively mature, with impetus having come largely from two directions:

1. *Telephony needs.* Lower bit rates would greatly increase the capabilities of the telephone network.

[1] The full name for the VODER was "Pedro the VODER," named after Dom Pedro, the Emperor of Brazil. He evidently said "My God, it talks!" after listening to Bell's telephone at the Philadelphia Centennial Exposition in 1876. This was enough for the Bell people to name the VODER after him. Perhaps it helps to be royalty.

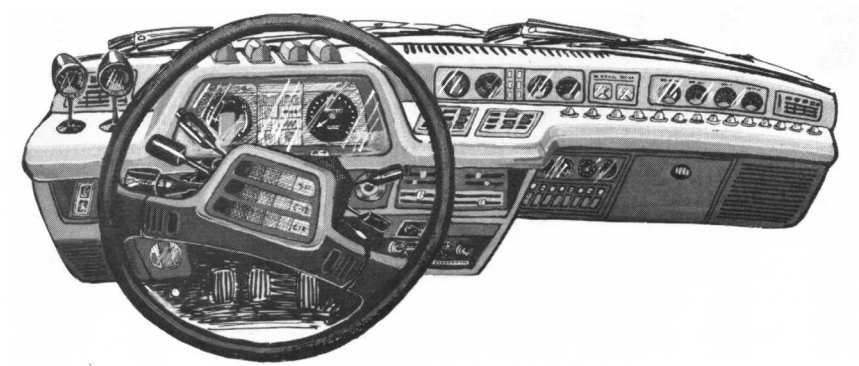
[2] "If the semiconductor industry had today a commercial million-transistor technology like VLSI, I'm not so sure it would know what to do with it."—Intel President Gordon Moore, 1979.

2. *Military applications.* Lower bit rates would permit more bits to be used for error correction, thus making communication more reliable.

Result: A plenitude of algorithms, just waiting for the hardware to apply them. Not to say that the major theoretical speech problems are all solved—far from it!—but simply that the algorithms are certainly ahead of the hardware capability to implement them.

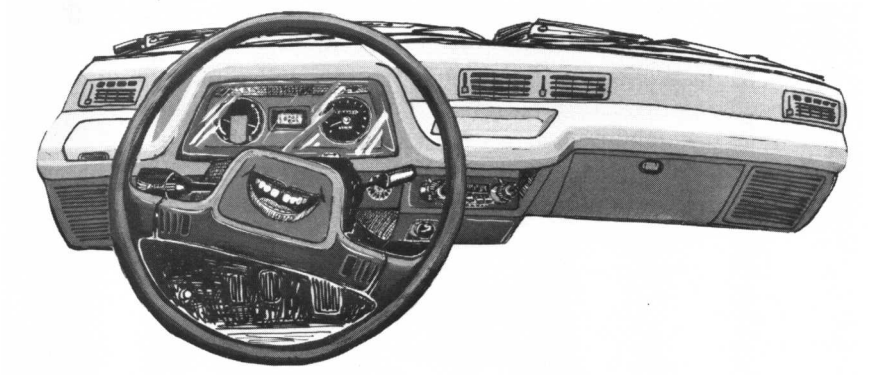
Speech seemed like a fruitful application for VLSI capability. And so it was. Speech products are beginning to proliferate. We will soon be seeing in common use talking toasters,[3] microwave ovens, scales, and toys. Synthetic speech will soon be available in most cars. These may seem to be mundane applications, but with the extension of information processing into the consumer domain, even these familiar devices might become unworkable without speech.

Speech can be another dimension in the operation of our tools. For example, as auto manufacturers begin to add sensors to determine different conditions, hazard monitoring by only visual means will become difficult if not dangerous.



[3] The talking toaster is a little joke of mine. I hope.

If there are too many lights and buzzers, we will be unable to take advantage of them quickly or without taking our eyes off the road. Instead, the car can simply *tell* us what is wrong.



New products are also aiding the handicapped: talking calculators, typewriters for the blind, and phonetic speech synthesizers for the vocally handicapped (such as cerebral palsy victims). These are applications for electronic speech synthesis that really have no substitute.

All of these products are becoming endowed with a small degree of intelligence. The advent of the inexpensive microprocessor has enabled manufacturers to put computers into our daily world. Synthetic speech will simply alter the way these computers communicate with us. This communication is critical to our relations with the modern world.

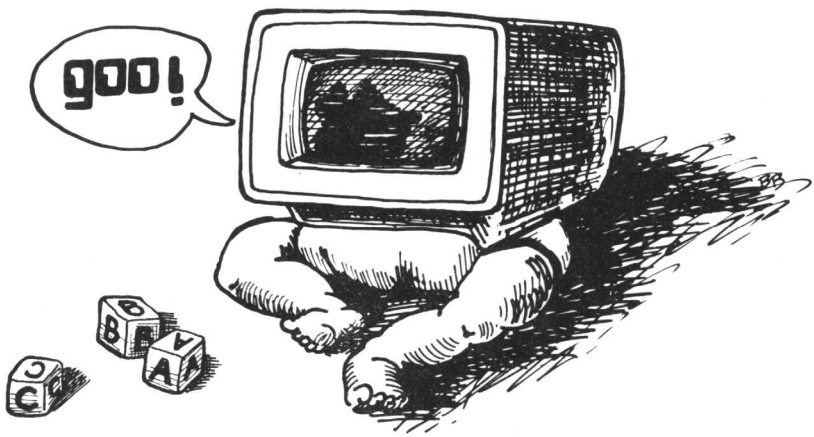
The nature of the man-machine interface has been characterized as dehumanizing. Consider the reputation of the talking (much less, intelligent) computer. Asimov speaks of the "Frankenstein complex," or the tendency to consider artificial entities as threats to our humanity. There is the fear that they will take over, turning our lives into a 2001 nightmare.

Like fire, modern technological advances are morally and ethically neutral. Perhaps some strange artificial intelligence (AI) development of the future will have an "evil" intent. In practice, however, the computer is largely our friend. It reduces the necessity for much physical and mental drudgery. We accept routinely the complicated mechanism of the modern automobile but approach with trepidation the idea of distant computers ruling our credit life. I don't wish to make fun of those fears. Perhaps they are legitimate. I do feel that the negative experiences men have had with fire, fossil fuels, or with computers are due almost entirely to the improper use of those technologies by fallible humans.

Recall Chaplin in the film *Modern Times*. For the sake of "efficiency," his employer fed Chaplin (a factory worker in the story), by means of an automatic worker-feeder. He was strapped in, and of course the machine malfunctioned, with comically messy results. In our modern times, we associate computers with the plethora of identification numbers with which we communicate to our corporate creditors. In this case, in fact in all cases, we perceive the computer through its input/output (I/O) interface. The current interfaces tend to be mildly annoying for some people. Yet this is enough to support the notion that computers are malevolent and dehumanizing. It *is* certainly dehumanizing to be addressed as a number rather than by name. However, these problems can be handled without rejecting the computers.

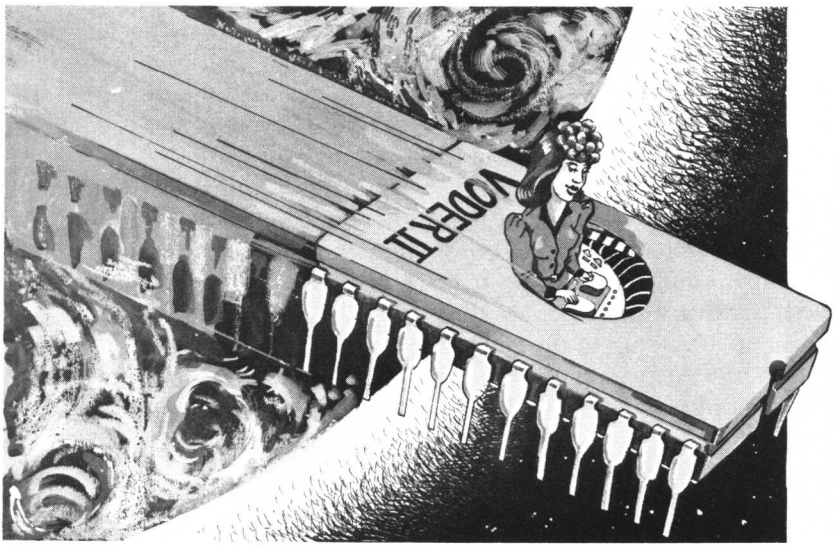
Speech I/O will go a long way toward a more humanly comfortable interaction with the ever-increasing number of computers around us. Up to now, men have struggled to bring themselves down to the level of the machine to communicate with it. The computer, along with the special-purpose hardware and software that are being developed, will inevitably bring the machine up to the human level of communication.

Speech processing can make the man-machine interaction natural and humanly based. Unless we destroy our own society with the violent technology that we have also developed, these machines will not be going away. Whatever one may think, it is clearly to our advantage to teach them our language.



Mechanical speech has a long and fascinating history. This century saw the introduction of the first connected speech synthesizer, the VODER. Recent advances in IC technology have permitted complicated systems similar to the VODER to be placed on a single piece of silicon, truly a "talking chip."

The VODER has come of age.



CHAPTER

1

SPEECH SYNTHESIS

1.1. INTRODUCTION

As the marvelous Electronic VODER Lady sails off into the sunset, we find ourselves poised on the doorstep of that mystifying world of modern talking machines. Swing open the door and enter. It won't be so bad. Really! There will be some concepts to be learned, a few stories to be told.

"Sure," you say, "talking machines would be fun to have around. They might even be useful. But why all the to-do? Why not just hook up some tape recorders and be done with it?"

There's a better answer than "because we need to keep speech engineers and scientists employed." That better answer *is* in fact money, but not just for speech professionals.

How much does it cost to make something talk? Tape recorders have increased in quality and decreased in cost over recent years. Yet they remain electromechanical beasts which wear down, break down, cost over $100, and are expensive to keep in good repair. In contrast, a fully electronic system composed of a few *chips* or integrated circuits can usually be marketed for under $20 in high volume, and is inexpensive to maintain. These prices will change, of course, but the ratio between the two should stay about the same. Electronic speech will remain cheaper than tape. This low cost makes electronic speech the obvious choice for high-volume applications, such as talking cars and computers.[1]

[1] Let's not forget the talking toaster. (On second thought, let's.)