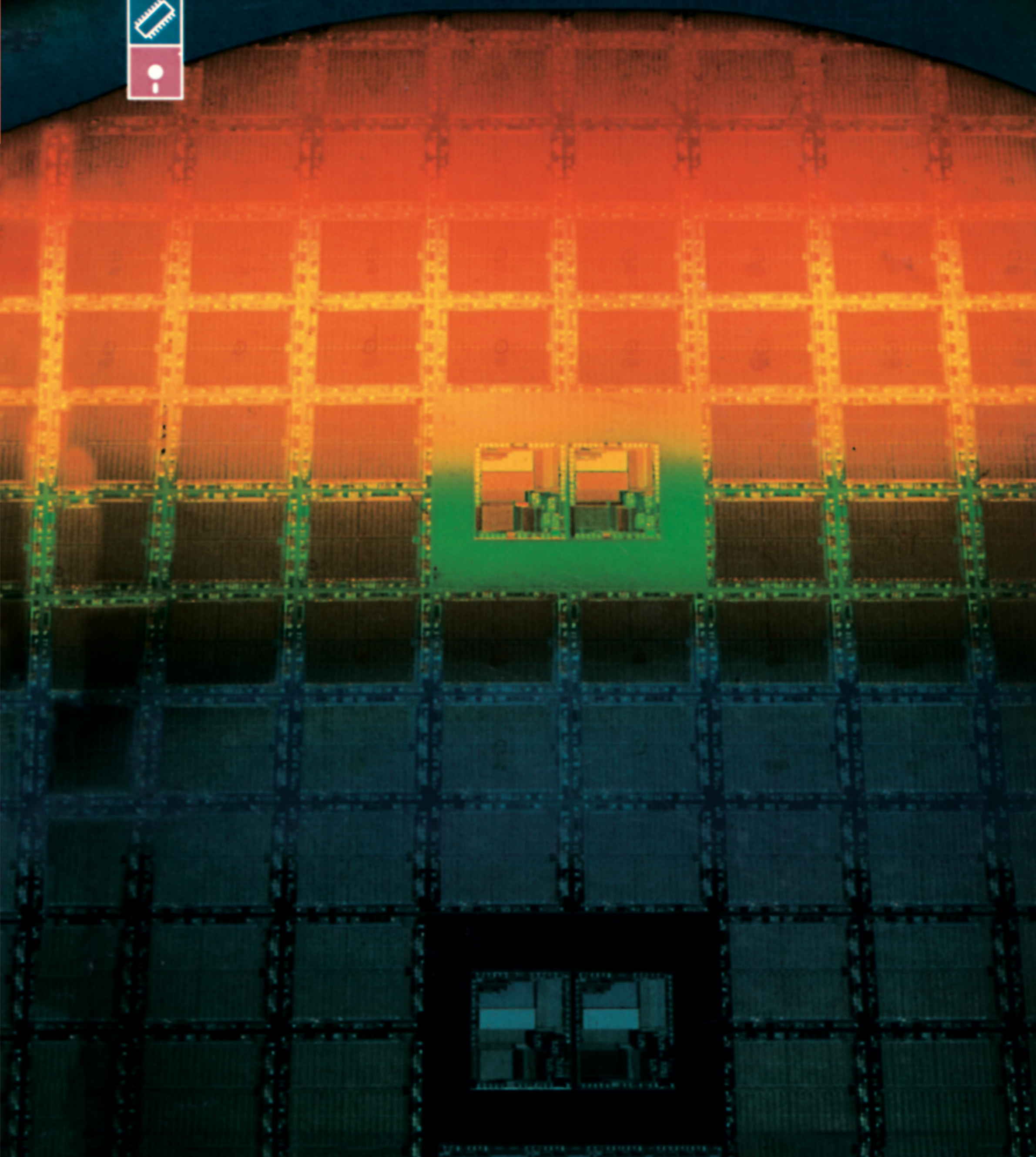
DESIGNING WITH SPEECH PROCESSING CHIPS

Ricardo Jiménez



Designing with Speech Processing Chips

Designing with Speech Processing Chips

Ricardo Jiménez

BESTNET Telecommunications

Calexico, California

ACADEMIC PRESS, INC.

Harcourt Brace Jovanovich, Publishers

San Diego New York Boston London

Sydney Tokyo Toronto

This book is printed on acid-free paper. ♾

ACADEMIC PRESS, INC.
San Diego, California 92101

United Kingdom Edition published by
Academic Press Limited
24–28 Oval Road, London NW1 7DX

Library of Congress Cataloging-in-Publication Data

Jiménez, Ricardo.
Designing with speech processing chips / Ricardo Jiménez.
p. cm.
ISBN 0-12-385348-6
1. Integrated circuits--Design and construction--Data processing.
2. Speech processing systems. 3. Computer-aided design. I. Title.
II. Title: Speech processing chips.
TK7874.J56 1991
621.39'9--dc20 90-49691
CIP

PRINTED IN THE UNITED STATES OF AMERICA
91 92 93 9 8 7 6 5 4 3 2 1

This book is dedicated
to my wife Patricia

This page intentionally left blank

Contents

CHAPTER 2
Experimenting with Speech Processors

CHAPTER 3
Analog Circuits

CHAPTER 4
Digital Circuits

CHAPTER 5
Test and Measurement Circuits

CHAPTER 6
Speech-Synthesized Burglar Alarms

CHAPTER 7
Voice Recognition Chips

Preface

The speech processing chip is a relatively new and complex device that appeared in the late 1970s. This book provides the theory and the basic design tools needed to utilize speech processing chips more effectively in electronic circuit design. It presents design examples for a wide range of real-world applications and information on interconnection of these components into functional equipment for instrumentation, data processing, inventory display, and control systems. Special emphasis is placed on those circuits with the most potential for future development, LSI and VLSI devices. Popular commercially available products are used throughout as illustrative examples, and the important characteristics of the devices are summarized for each functional category.

The book goes far beyond the presentation of block diagrams, microcontroller architecture, and software programming. It shows a step-by-step development of hardware and software, how to combine them most effectively, and how to interface speech processors with input devices, such as sensors or data sources, and output devices, such as relay actuators, thyristors, or display devices.

In this book, practicing design and system engineers, technicians, engineering students, and other interested readers will find a comprehensive overview of the entire topic of speech processing chips. The book also describes popular, commercially available circuits for each functional category presented and discusses specific applications in sufficient depth to interest the experienced designer. Engineering students will be able to follow the book if they have been exposed to courses on circuit design theory, logical circuits, and integrated circuits.

As with my other publications, I wanted this to be a book that was organized and used from a practical viewpoint. Throughout the book emphasis is placed on using the popular CMOS and HCMOS ICs of each functional category as illustrative examples.

Speech processing chips for electronic and other applications are available at low cost. The progress made in the manufacture and supply of these chips expands enormously the opportunities to design and build highly effective equipment and systems. This book shows how to make use of these developments. The reader is shown step-by-step how to build both simple and sophisticated projects with all the necessary details. Many proven examples are included throughout the book for industrial, laboratory, health care, and home use.

The information needed to follow the many design examples in this book is given in a simple, direct manner with supporting flowcharts and tables. Once this know-how is acquired, you will be able to build systems with artificial voice with less effort and less time than ever before.

The book is divided into seven chapters. Chapter 1 introduces the different speech processing techniques, describes how the basic speech processor integrated circuit works, and presents IC pin comparisons of the different packages. It also includes the basic datasheet for the device.

Chapter 2 explains how a speech processor can be used in applications for which it was not originally intended—how this basically digital device can be used as a variety of different logic devices. Chapter 3 provides help understanding analog-to-digital converter families and their respective advantages and limitations. After reading these three chapters, you should come away with a fundamental understanding of what a speech processor is and how to use it.

To write information into the specific controller and then announce it, interface devices or systems are required. Chapter 4 shows how such interface devices can be built with minimal effort and at a cost that is often a small fraction of the price of commercial products.

Chapter 5 presents a wide selection of test and measurement circuits that also can be interfaced with a specific application by the user or designer. These circuits are widely used in data acquisition systems. Chapter 6 presents different kinds of burglar alarms varying from simple designs to fault-tolerant systems where failures are critical and not acceptable.

Chapter 7 covers voice recognition techniques as well as devices now available. Some applications for control systems are also considered.

Acknowledgments

I owe thanks to several people who helped me complete this book.

I would like to thank all the publishers who gave permission to include material that originally appeared elsewhere. Footnotes to selected articles cite these original sources.

I am grateful to those instructors from other institutions who reviewed the manuscript for Academic Press, Inc., and who made many constructive suggestions.

Thanks also to my research students of the courses "Logical Circuits II," "Sequential Systems," and "Instrumentation" who suffered through the building and testing process of many of the circuits presented here and in particular to Ruben Rios, Francisco Meza, Jose J. Lara, Ramiro Jacinto, Ruben Avalos, Francisco J. Mendoza, Alvaro E. Salgado, Julio Garcia, Rigoberto Hernandez, and Fabian Muro.

Thanks in particular goes to the Dean of the Technological Institute of Mexicali, Ramon A. Heredia.

I am also grateful to Marcos Silva, Roberto Dewar, and Dr. Fidel Diaz for proofreading the manuscript to make it more understandable. Of course, the patience and love of my wife Patricia was required throughout.

Ricardo Jiménez, E. E.

This page intentionally left blank

CHAPTER 1

Speech Processing Chips

1.1 Introduction to Voice Synthesis Digital ICs

Circuits with artificial voice offer a new dimension of sophistication to almost any electrical or electronic modern system. Traditionally, magnetic tape recording has been used in applications requiring speech announcements, for example, telephone announcement systems; a system of this type is costly because it requires a large number of tapes for different messages. It will not let you create mixed messages for different situations. Consider the case of a public service telephone that tells you the time. This device will need 60 different tapes for each hour, not to mention the number of tapes required for a complete 24-hour day. In contrast, a telephone system that uses artificial voice stored in digital memories can create different messages by pulling up the different words required to create a specific message. This system requires only a few chips that will do the work of a large number of tapes.

Let us now consider a talking voltmeter in the range of 0 to 5 V with a resolution of 0.1 V. Here, we will need 50 different messages corresponding to the 50 possible voltage readings. It would be costly and time-consuming to develop a tape mechanism for this project.

In the past, speech systems were treated as data acquisition circuits, in which a voice waveform was treated like any other fluctuating voltage input: The circuit recorded the waveform by periodically taking a sample of the signal's voltage through an analog-to-digital (A/D) converter and storing it as a binary value. (The number of samples needed per second depends upon the frequency of the input signal.) These digital speech signals were stored as pulse code modulation (PCM) in semiconductor memories. Once the samples were stored in RAM or ROM, the circuit could recreate the original waveform by sequentially sending the stored values to a D/A converter at the same rate as the original sampling. One second of digital voice required from 4 to 32 Kbytes of memory. If the amount of data stored was reduced (compressed) using a known principle, the restoration of the original sound was called "synthesis."

The synthesis technique provides a dramatic reduction in the amount of memory required for one second of speech. Memory requirements vary from 400 to 2,000 bits per second, depending on the desired speech attributes and overall quality. A bad reproduction will sound unnatural or unintelligible. A speech signal is highly redundant and predictable, and by coding only the slowly varying coefficients of speech or by dramatic compression of digitized speech, significant bandwidth reductions in the digitized signal can be obtained. The synthesizer technique becomes practical when it is developed with VLSI semiconductor technology.

Today, applications for voice synthesis are endless. The following are some of them: telecommunications; consumer appliances; automotive; counters; consumer products; instrumentation; teaching aids; clocks; language translation; annunciators; voice interactive computer terminals; nautical and aeronautical instrumentation annunciators; voice back units for banking, weather, and time announcements; elevators; trains; subway systems; toys and games; warning systems for fire and police emergency.

In the area of instrumentation, a speech synthesizer is a very important tool. When a failure is presented in the system being monitored, the speech synthesizer will immediately start reading the procedures contained in the manual in order to indicate to the operator how to correct the specific failure. Great benefits are obtained if a speech synthesizer is installed in power stations, nuclear plants, or places where the user must monitor a myriad of controls. Here the speech synthesizer augments the operator's ability to respond rapidly and correctly when a process has extended its normal limits.

In industry, a speech synthesizer can be used to augment productivity by giving spoken messages on how to assemble specific products, thereby freeing a user for other tasks. Here, failure to follow precise directions could lead to the destruction of equipment or injury to personnel.

The use of vocal warnings on automobiles has been spreading since the early 1980s to remind the driver of the electrical or mechanical situation of the vehicle. The same features are also applied to airplanes where the synthetic speech guides the pilot with directions such as "slow down," "climb," or other appropriate instructions.

The pace of the speech processing field is so rapid that some systems now under development are excluded from this book. The emphasis of this book is on designing with the systems now available.

1.2 Synthesis Techniques

The basic phonological element of speech is the phoneme, which is the name given to a group of similar sounds in language. A phoneme is acoustically different depending upon its position within a word. Each of these positional variants is an allophone of the same phoneme. Phoneme reproduction is a basic element in any speech synthesizer. The method of "allophone speech synthesis" is used to create words or phrases where the user has to think in terms of sounds, not letters. With this technique you can synthesize an unlimited vocabulary by using allophones and silences in the appropriate sequence. Phonemes, together with speaker inflection and volume, are the fundamental building blocks of speech.

The American English language consists of approximately 38 to 40 phonemes: 14 to 16 vowel sounds and 24 consonant sounds. For example, the initial K sound used in words like "comb" sounds slightly different from the Ks in words like "can't." These small variations are due to the vowel which follows them, in this case, "o" and "a." Each phoneme is generated with either a voiced sound, as in "eye" or an unvoiced sound like the "sh" in "shy." There are also allophones classified as resonants, voiced fricatives, voiceless fricatives, voiced stops, voiceless stops, affricates, and nasals.

Voice synthesis methods are divided into three major types: waveform encoding, parametric synthesis, and synthesis by rule. Each method is explained below.

Waveform Coding Methods

This type of voice synthesis includes differential pulse code modulation (DPCM), adaptive delta modulation (ADM), and adaptive differential PCM (ADPCM). The original sound wave amplitude is sampled at fixed intervals, digitized, and the volume of data is then reduced on the basis of the synthesis principles.

Parametric Synthesis Methods

Characteristic information included in voice waveforms is extracted as parameters for synthesizing purposes. The partial autocorrelation (PARCOR) method is a typical example. In this method, models of the human vocalization mechanism are used. Voiced and voiceless consonant sounds are discriminated, and voiced sound pitch and amplitude data are extracted together with filter characteristics of the vocal tract. Voice synthesis is then obtained by passing these data to hardware consisting of digital filter circuits.

Synthesis by Rule Method

In this synthesis method, groups of phonemes expressed by small quantities of data are skillfully linked together to reproduce any desired words or phrases,