A Comparison of Speaker Identification Results Using Features Based on Cepstrum and Fourier–Bessel Expansion

K. Gopalan, *Senior Member, IEEE,* Timothy R. Anderson, *Member, IEEE,* and Edward J. Cupples, *Member, IEEE*

***Abstract*—A compact representation of speech is possible using Bessel functions because of the similarity between voiced speech and the Bessel functions. Both voiced speech and the Bessel functions exhibit quasiperiodicity and decaying amplitude with time. This paper presents the results of speaker identification experiments using features obtained from 1) the Fourier–Bessel expansion and 2) the cepstral representation of speech frames. Identification scores of 65% and 76% were achieved using features based on $J_1(t)$ expansion of air-to-ground speech transmission databases of 143 and 1054 test utterances, respectively. The corresponding scores for the two databases using cepstral coefficients of a comparable size were 80% and 88%. A comparison of the two sets of features indicates that $J_1(t)$ can be used to model the hearing perception much like the mel cepstral coefficients.**

***Index Terms*— Bessel functions, cepstral features, Fourier–Bessel expansion, speaker identification.**

I. Introduction

NONSINUSOIDAL basis functions with quasiperiodic behavior possess more structural similarity to speech signals than the sinusoids of the Fourier transform. Hence, the representation of speech signals using these basis functions may result in fewer components than using the sinusoids. Based on this observation, aperiodic nonsinusoidal functions such as exponentially modulated sinusoids and Bessel functions of the first kind have been attempted for speech analysis [1]–[5]. In this paper, we consider a set of speech features obtained using the first order Bessel functions in a speaker identification application. We show that the Bessel functions can be used to represent speech waveforms and that the features derived from the representation are comparable to a set of cepstral domain features in terms of speech perception and speaker identification.

II. Signal Representation Using Bessel Functions

An arbitrary function $x(t)$ in the interval $0 < t < a$ is represented using the Bessel function $J_1(t)$ in the Fourier–Bessel (FB) expansion given by [6]

$$x(t) = \sum_{m=1}^{\infty} C_m J_1\left(\frac{x_m}{a}\,t\right) \tag{1}$$

where x_m, $m = 1, 2, 3, \cdots$ are the roots of $J_1(t) = 0$.

Using the orthogonality of the set $\{J_1[(x_m/a)t]\}$, the FB coefficients of expansion are determined from

$$C_m = \frac{2}{a^2 J_0(x_m)^2} \int_0^a t x(t) J_1\left(\frac{x_m}{a}\,t\right) dt. \tag{2}$$

As with the Fourier series, the above FB coefficients $\{C_m\}$ are unique for a given $x(t)$. Unlike the sinusoidal basis functions in the Fourier series, however, the Bessel functions are aperiodic and decay within the range a. The quasiperiodic set $\{J_1([x_m/a]t)\}$ has structural similarity to short-time speech segments, with the interval between successive zero-crossings increasing slowly with time and approaching π in the limit. Therefore, an efficient representation of speech segments in terms of the orthogonal set can be obtained using the coefficients C_m.

III. Fourier–Bessel Representation of Speech Waveforms

A speech signal is represented using the discretized version of (2) on short intervals of speech. The coefficients $\{C_m, m = 1, 2, \cdots\}$ are obtained for each of the segmented and windowed frames of sampled speech. Although (1) is an infinite series, it approaches $x(t)$ rapidly with a large but finite number of coefficients. An approximate representation of the signal can, therefore, be obtained with only a finite number of terms.

The effectiveness of the Bessel function representation of speech was studied using different number of coefficients in the summation given in (1) [4], [5]. The database used for this experiment consisted of 54 utterances of the sentence, “We were away a year ago,” spoken by three males and two females. Two sets of recordings of speech signals were obtained at the rate of 11 000 samples/s. Beginning and end of speech in each utterance were detected using amplitude thresholds. After experimenting with different intervals for speech frames, with and without overlap, it was found that using 200 samples/frame with 100 sample overlap was computationally optimum. Each frame of speech data was then multiplied by a 200-point

Manuscript received February 3, 1997; revised July 5, 1998. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Joseph Picone.

K. Gopalan is with Purdue University-Calumet, Hammond, IN 46323 USA (e-mail: gopalan@calumet.purdue.edu).

T. R. Anderson is with the Air Force Research Laboratory, Wright-Patterson AFB, OH 45433 USA.

E. J. Cupples is with the Air Force Research Laboratory, Rome, NY 13441 USA.

Publisher Item Identifier S 1063-6676(99)02734-0.

Hamming window to preserve the spectral characteristics of the signal.

When 100–150 coefficients (i.e., $C_m, m = 1, 2, \cdots, M$, where M is between 100 and 150) were used in the summation to reconstruct the speech signal, all the characteristics of the original speech, including the identity of the speaker, were observed to be present. The quality of the synthesized speech was good enough for listeners, who were unfamiliar with the message, to clearly understand the speech and distinguish the speakers. In addition to the listening test, the speech quality was also confirmed by the spectrogram of the synthesized speech which was close to that of the original speech.

The speech quality began to deteriorate for M (the upper limit of m) between 30 and 50; the message quality was still acceptable, however. At the low end of M, particularly at 20, the identity or the gender of the speaker was difficult to determine, although the message was still discernible with background noise. Synthesized speech using only the first ten terms had very little message quality with only the low frequency components present.

The correspondence between the finite number M of terms used in the reconstruction in (1) and spectrum of the resulting signal is established by the bandwidth $\omega_{\max}$ of the reconstructed signal. Since the frequency spectrum of $J_1[(x_m/a)t]$ is bandlimited to $\omega = x_m/a$ (Fig. 1) for M terms in the summation, the signal approximation given by

$$x(t) \approx \sum_{m=1}^{M} C_m J_1\left(\frac{x_m}{a}\, t\right) \tag{3}$$

has a maximum bandwidth of $\omega_{\max} = x_M/a$, where x_M is the Mth root of $J_1(t) = 0$.

Hence, using the first M terms in the summation in (1) limits the spectrum of the synthesized speech to x_M/a. Consequently, to retain all the frequency components up to 3 kHz in the representation, for example, M must be chosen such that $x_M \approx 2\pi(3000)a$. For the original speech sampled at 11 000 samples/s, using 200 samples per frame gives the range a value of 18.2 ms. Hence, $x_M = 342.72$. Since the 109th root of $J_1(t)$ is close to this value (at 343.218), the finite summation must have up to 109 terms to retain the spectral components up to 3 kHz. For $M < 109$, the synthesized speech loses the high frequency information. At $M = 50$, for example, the spectral energy of the reconstituted speech is limited to below 1500 Hz, which may be below the second formant for certain speakers and/or phonemes; it certainly cannot have the original speech quality of fricatives. The message content may still be preserved, albeit with noise, depending on how large M is; hence, a lower value of M (below 50) may be acceptable for storage and transmission of speech.

Improvement in speech quality, particularly for fricatives, requires a certain amount of high-frequency information that corresponds to higher value of M. If the high-frequency components are predominant in any frame in the original speech waveform, they manifest in the form of large coefficients $|C_m|$ for large index m. Therefore, by selectively choosing the coefficients in each frame based on their relative magnitudes, spectral energy at both low and high frequencies may be retained in the synthesized speech. Thus, with a slight increase in the processing time (arising from sorting of $|C_m|$ in each frame), higher speech quality can be achieved at lower number of coefficients. This assertion was verified using the top 50, 30, 20, and ten coefficients. It was found that a minimum of 20 sorted coefficients were needed for an acceptable quality of speech without missing the identity of the speaker. This compares favorably with the requirement of more than 30 unsorted coefficients for low noise in synthesized speech.

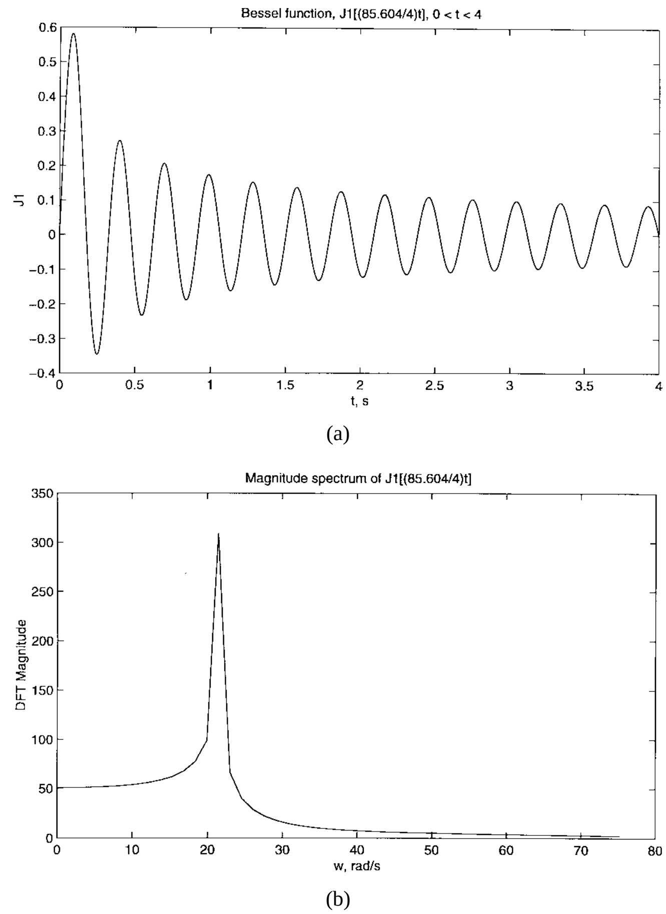


Fig. 1. Bandlimiting of $J_1(t)$: (a) time function $J_1([x_m/a]t)$, $x_m = 85.604$ and $a = 4$ and (b) discrete Fourier transform magnitude of $J_1([x_m/a]t)$, $x_m = 85.604$ and $a = 4$.

Based on the above results of speech representation using $\{J_1(t)\}$ as the basis function set, it is clear that a set of a finite number of coefficients in the representation can be used to form a feature vector to characterize speech and speaker.

IV. Features Using Cepstral Domain Parameters

Motivated by the observation that the human auditory system perceives information based on the energy in a band of frequencies rather than that at a single frequency, cepstral domain parameters are commonly used in speech recognition and speaker identification systems [7], [9]. The cepstral parameters of a speech signal $x(n)$ are obtained as the inverse discrete Fourier transform (IDFT) of the output of a set of critical band filters whose input is the log magnitude of the discrete Fourier transform (DFT) of $x(n)$, as shown in Fig. 2. Although the critical band filters are allpass filters for pitch

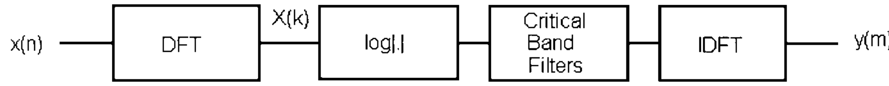


Fig. 2. Cepstral domain feature generation.

extraction, for example, overlapping bands of triangular filters are used for speech and speaker recognition applications [7].

For the present study, two sets of cepstral parameters were considered for speaker identification. The database ("Greenflag database") consisted of air-to-ground pulse code-modulated transmissions obtained at a sampling rate of 8000 samples/s. A frame length of 200 samples (25 ms duration) with 100 sample overlap was chosen for cepstral processing. With a 1024-point DFT of each frame, the frequency spectrum was obtained at a resolution of approximately 8 Hz/point. With the database consisting of all male voices, the first frequency of interest was chosen at approximately 100 Hz, a value in the vicinity of the fundamental frequency for male speech. Ten frequencies starting at 100 Hz were chosen on a linear scale with a spacing of 100 Hz. Above 1000 Hz, nine frequencies on mel frequency scale were used to cover the range from 1100–3500 Hz. Outputs of filters centered at the chosen frequencies were formed from the magnitude spectrum of each frame. The filters for the linearly distributed center frequencies had a constant bandwidth of approximately 100 Hz while those for the mel scale had logarithmically increasing bandwidths with each filter spectrum overlapping the adjacent filter spectra. The 19 cepstral coefficients were obtained as the IDFT of the log energy at the selected indices. The 19 log energy values (prior to the IDFT operation) formed a second feature vector.

V. Identification Results

The following sets of features were evaluated for two databases, namely, the Greenflag database and the NATO database [8]. The Greenflag database consisted of a total of $173 + 41 = 214$ utterances from 41 speakers (pilots), which are transmitted from aircraft to ground at a sampling rate of 8000/s. An utterance from each speaker was used as a reference utterance. All parameters were evaluated using a frame length of 200 samples with 100 sample overlap.

A. FB-Based Features

To reduce the feature vector size, a selected set of FB coefficients, as described in Section III, may be used without loss of information. Although the coefficients may be selected based on their relative amplitudes or their frequency content [4], [5], a set of 15 coefficients that covered a bandpass frequency range was chosen starting with the indices given by $m = 10, 20, 35, 50, 60, 70, 80, 90, 100, 115, 130, 145, 160, 175,$ and 190. At each of these 15 indices, five successive coefficients were computed. These coefficients (15×5) represent a starting frequency of 205 Hz (at $m = 10$) and go up to approximately 3905 Hz (at $m = 195$). We may note that the frequencies at which the maximum spectral energies occur for the basis functions, $J_1([x_m/a]t)$, are somewhat analogous to the mel scale of frequency values chosen for the cepstral domain feature extraction. To reduce the feature vector size further, an energy measure in a narrow band of frequencies in the vicinity of each selected index was obtained. The magnitude of the basis signal amplitude at each of the five indices was added to get the energy measure as

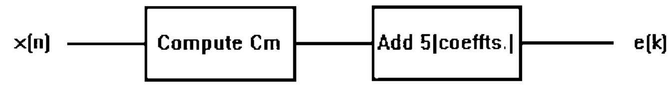


Fig. 3. Energy measure feature using FB coefficients.

$$e(i) = \sum_{m=k}^{k+5} |C_m| \tag{4}$$

for $i = 1, 2, \cdots, 15$ and $k = 10, 20, 35, 50, 60, 70, 80, 90, 100, 115, 130, 145, 160, 175,$ and 190, respectively, as shown in Fig. 3.

We note that the feature element $e(k)$ is analogous to the log energy output of the critical band filter in the cepstral domain representation. Thus the FB-based feature vector models the perceptual hearing at selected frequencies. By adding a fixed number of coefficients (five, for example, here), the bandwidth at each selected frequency is kept almost constant with the index m. At $i = 1$ ($k = 10$ and $m =$ 10–14), for example, $e(1)$ represents approximately the energy in the bandwidth of $f_1 = x_{10}/2\pi a = 204.93$ Hz, to $f_2 = x_{14}/2\pi a = 284.95$ Hz, or 80 Hz. At $i = 10\,(k = 115$ and $m =$ 115–119), $e(10)$ represents the energy from $f_1 = 2305$ Hz to $f_2 = 2385$ Hz. The energy measure $e(i)$ as given in (4) was used in constructing 15-element FB feature vectors for the Greenflag database. In addition, 20-element FB feature vectors using (4) for $i = 1, 2, \cdots, 20$ and $k = 5, 15, 25, 35, 45, 65, \cdots, 185, 195,$ were also computed. With the starting coefficient index of 5 and the final index of 195, the 20-element vectors covered wider range of frequencies than the 15-element vectors at both the lower and the upper end.

B. Cepstral Features

Cepstral features of 19 elements each were evaluated for each frame as explained in Section IV using a 1024-point DFT. The critical band triangular filters had a bandwidth of 101.5 Hz (13 points in the DFT domain) for center frequencies of up to 1016 Hz; the bandwidth increased from 117.2 Hz (DFT index 15) at 1148 Hz, to 617.2 Hz (DFT index 79) at 3469 Hz. To represent the spectral changes from frame to frame, delta cepstra were also computed from the 19-element cepstral features for 1) three frame difference and 2) five frame difference [9].

Twenty cepstral coefficients and 20 log energy values were also computed for each frame with an increased frequency resolution of 2048-point DFT and at slightly larger bandwidths for the critical band filters (see Table I).

TABLE I
IDENTIFICATION SCORES FOR THE GREENFLAG DATABASE USING CEPSTRAL AND FB-BASED FEATURES

No	Feature	Score No. correct/Total	Score in percent
1	19 Cepstral Coefficients [a]	112/173	64.7
2	20 Cepstral Coefficients [b]	139/173	80.4
3	20 Log energy values [b]	132/173	76.3
4	15 FB Energy values at selected indices [c]	128/173	74.0
5	20 FB Energy values at selected indices [d]	131/173	75.6
6	19 Differential energy values at selected indices - index diff.	122/173	70.5

[a] The cepstral coefficients were obtained with critical band filters centered at the frequencies of 102 Hz, 203 Hz, 305 Hz, 406 Hz, 508 Hz, 609 Hz, 711 Hz, 813 Hz, 914 Hz, 1016 Hz, 1148 Hz, 1320 Hz, 1516 Hz, 1734 Hz, 1992 Hz, 2289 Hz, 2633 Hz, 3023 Hz, and 3469 Hz.

[b] The 20 cepstral coefficients and the log energy were evaluated using 2048-point DFT and at critical band filters and bandwidths different from those of previous computations. The approximate center frequencies of the filters are: 116 Hz, 198 Hz, 300 Hz, 398 Hz, 500 Hz, 600 Hz, 700 Hz, 800 Hz, 898 Hz, 1000 Hz, 1148 Hz, 1316 Hz, 1508 Hz, 1738 Hz, 1996 Hz, 2288 Hz, 2628 Hz, 3020 Hz, 3460 Hz, and 3672 Hz.

[c] Magnitudes of five coefficients starting at the following 15 indices are added to obtain the feature vector: 10, 20, 35, 50, 60, 70 80, 90, 100, 115, 130, 145, 160, 175, and 190.

[d] Magnitudes of five coefficients starting at indices 5, 15, 25, ... 185, 195 are added to obtain the feature vector of size 20.

The above sets of features were tested using a commercial vector quantizer-based classifier available from Entropic Research Laboratory [10]. With the same classifier used on all the features and for all the databases, the effectiveness of each feature may be compared in terms of computational complexity, and identification scores. We note that the Greenflag database used is noisy with bursts of engine noise and microphone clicks; also, many of the test utterances are very short in duration (a few hundred milliseconds) compared with the corresponding training utterances in the database. With no endpoint detection, therefore, the burden of discriminating each speaker regardless of the utterance duration and the quality of speech is primarily on the features. The classifier performance is not likely to adversely affect the identification score. Table I lists the scores for the sets of features discussed.

The higher scores observed using 20 cepstral parameters are due more to the increased bandwidth of each critical band filter than to the increased frequency resolution. This assertion was verified using 2048-point DFT but with the same bandwidth as with the 1024-point DFT. The score in this case was only slightly higher at 67%. The reasonably high score of 76.3% using the log energy values indicates that the feature extraction may be speeded up by avoiding the inverse DFT operation; for higher scores, however, complete cepstral processing as indicated in Fig. 2 is needed. The inclusion, or the stand-alone usage, of the differential and delta cepstra did not result in improved scores.

The choice of the coefficient indices in the case of the FB-based feature vectors, as with the cepstrum, determined the feature efficiency more than the size of the feature. The 20-parameter feature vector covered approximately the same center frequencies and bandwidths as in the case of the 20-point cepstrum used. The inclusion of the low frequency components corresponding to $m = 5$–9, covering approximately 100–200 Hz, raised the score slightly.

The differential energy in the FB domain was obtained as the difference between successive energy elements, i.e.,

$$d(k) = e(k+1) - e(k), \qquad k = 1, 2, \cdots, 19.$$

This differential energy feature is analogous to the energy in a differential band of frequencies. (Because of the slight increase in the identification score using the 20-element FB energy vectors over that using the 15-element vectors, only the former was employed in constructing the differential energy

TABLE II
IDENTIFICATION SCORES FOR THE NATO DATABASE USING CEPSTRAL AND FB-BASED FEATURES

No	Feature	Score No. correct/Total	Score in percent
1	20 Cepstral Coeffts.	928/1054	88.0
2	20 Log energy values	885/1054	84.0
3	20 FB Energy values at selected indices -- 400 samples/frame	582/1054	55.2
4	20 FB Energy values at selected indices -- 200 samples/frame	683/1054	64.8
5	19 Differential energy values at selected indices - index diff.	651/1054	61.2

features.) Since the coefficient at index M is approximately proportional to the spectral amplitude at frequency $\omega_{\max} = x_M/a$, the differential energy measure $d(k)$ is approximately equivalent to the spectral energy in the band $\omega_{k+1} - \omega_k = (x_{k+1}/a) - (x_k/a)$. Thus $d(k)$ resembles the bandpass filtered signal energy. Clearly, the choice of the coefficients in forming $e(k)$, and hence $d(k)$, again, determines the effectiveness of the feature vector $d(k)$. For the given set of coefficients forming $e(k)$, the identification score of 70.5% using $d(k)$ is comparable to the score at 75.6% using direct $e(k)$.

C. The NATO Database

This database consisted of short utterances from nine European speakers (pilots), transmitted from aircraft to ground at the rate of 16 000 samples/s. From a total of 1117 transmissions, seven utterances from each speaker were concatenated to form reference utterances. The remaining 1054 transmissions were used in speaker identification. Table II shows the identification scores using cepstral and FB-based features.

Because of the higher sampling rate of the data, 512 samples were used in each frame, with 256 sample overlap, for cepstral parameters. With 2048-point DFT, the 20-element feature vectors were obtained at the frequencies of 233, 398, 600, 796, 1000, 1200, 1400, 1600, 1796, 2000, 2296, 2632, 3016, 3476, 3992, 4576, 5256, 6040, 6920, and 7344 Hz. The cepstral features have clearly performed better with significantly higher scores than the FB-based features.

With 400 samples/frame, the FB-based 20-element energy feature vectors covered the frequencies from approximately 105–3905 Hz using the indices 5, 15, 25, 35, 45, 65, $\cdots$, 185, 195. While this range of frequencies is sufficient to represent speech, the cepstral domain features cover a much wider range of 233–7344 Hz. With the high recognition score using this range of frequencies in the cepstral domain features, it was clear that for this database, energy at frequencies above 4000 Hz must be included in the FB features for identification. Therefore, the frame length for use with the FB transform was changed to 200 samples. At this frame length, the range became 12.5 ms and the frequency coverage changed to 210–7810 Hz. The last two rows show a significant increase in the identification score with the increased frequency coverage. We note that by using only five coefficients in the FB energy feature vector, the band of frequencies covered in each element, $e(k)$, $k = 1, 2, \cdots, 20$, is constant at approximately 210 Hz. The cepstral features, however, covered an increasing band of frequencies starting at 63 Hz (for the center frequency of 233 Hz) and ending at 1250 Hz (for the center frequency of 7344 Hz).

VI. DISCUSSION

The FB representation of speech is clearly an alternative to the traditional Fourier transform-based representation for speech reconstruction and speaker identification applications. Because of its nonstationary behavior, similar to speech sounds within a pitch interval, the representation in terms of the FB coefficients may be more compact than that using the stationary sinusoids. The choice and the number of coefficients, however, may depend on the speech sounds—vowels requiring fewer and lower indexed coefficients while fricatives requiring larger and/or higher indexed coefficients, for example.

Since the FB coefficient at index m approximately represents the spectral amplitude at frequency $\omega_{\max} = x_M/a$, energy in a band of frequencies is closely related to the magnitude of FB coefficients with adjacent indices. Computationally, however, Fourier domain calculations have the advantage of fast and in-place algorithms, which are not available for the FB domain calculations at orders of n above zero for $J_n(t)$.

For completion of the comparison with the mel distributed cepstral coefficients based on the energy in a critical band of frequencies, an increasing number of coefficients may be considered in forming the feature vector $e(k)$ in (3). Although this increases the computational effort, it may yield higher speaker identification scores that are comparable to those using cepstral features.

VII. CONCLUSION

A method of speech signal representation and feature extraction using first-order Bessel functions as basis functions has been described. It has been shown that the features obtained from the FB expansion of speech are comparable to the cepstral features in representing the spectral energy. Therefore, FB-based features may be used as an alternative to the spectral representation of speech signals. Because of the analogous nature of the FB-based features to the cepstral domain features, both feature sets yield comparable speaker identification scores. With a judicious choice of the coefficients in forming the features, the FB expansion can be used efficiently in speech analysis-synthesis and speaker identification applications.

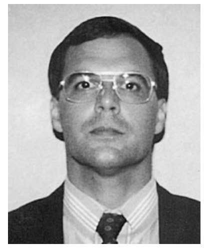
REFERENCES

[1] L. Dolansky, "Choice of base signals in speech signal analysis," *IRE Trans. Audio Electroacoust.*, vol. AU-8, pp. 221–229, 1960.
[2] H. J. Manley, "Analysis-synthesis of connected speech in terms of orthogonalized exponentially damped sinusoids," *J. Acoust. Soc. Amer.*, vol. 35, pp. 464–474, Apr. 1963.

[3] C. S. Chen, K. Gopalan, and P. Mitra, "Speech signal analysis and synthesis via Fourier-Bessel representation," in *Proc. ICASSP*, Tampa, FL, Mar. 1985, pp. 497–500.
[4] K. Gopalan, "Speaker identification using Bessel function expansion of speech signals," Final Rep. Summer Faculty Res. Program, Armstrong Lab., AFOSR, Aug. 1993.
[5] K. Gopalan and T. R. Anderson, "Speech processing using Bessel functions," in *Proc. Symp. Intelligent Systems in Communications and Power*, Mayaguez, PR, Feb. 1994, pp. 255–259.
[6] I. H. Sneddon, *Fourier Transforms.* New York: McGraw-Hill, 1951.
[7] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 28, pp. 357–366, Aug. 1980.
[8] K. Gopalan, "Speaker identification and analysis of stressed speech," Final Rep. Summer Faculty Res. Program, Rome Lab., AFOSR, July 1996.
[9] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals.* New York: Macmillan, 1993.
[10] ESPS Programs, Entropic Res. Lab., Washington, DC, 1996.

K. Gopalan (S'80–M'82–SM'94) received the B.E. degree in electrical engineering from P.S.G. College of Technology, University of Madras, Coimbatore, India, in 1971, the M.Tech. degree in electrical engineering from the Indian Institute of Technology (IITK), Kanpur, India, in 1974, and the Ph.D. degree in engineering from the University of Akron, Akron, OH, in 1983.

From 1974 to 1979, he was employed at IITK, first as an Instrumentation Engineer and later as a Research Engineer. While pursuing graduate studies at the University of Akron, he held the positions of Teaching Assistant, Research Assistant, and Instructor. From 1983 to 1985, he was Assistant Professor of electrical engineering at Lafayette College, Easton, PA. Since 1985, he has been with the Department of Engineering at Purdue University-Calumet, Hammond, IN, holding the position of Professor of electrical engineering. From 1987 to 1995, he conducted research in the areas of signal and image processing for nondestructive evaluation of advanced materials at Argonne National Laboratory, Argonne, IL, first as a summer faculty research participant and later as a consultant. In addition, he has been a summer faculty research associate at Wright-Patterson Air Force Base, OH (1993), and Rome Laboratory (1996 and 1998), NY, and worked in the areas of speaker identification and analysis of speech under stress. He is the author of the textbook *Introduction to Digital Microelectronic Circuits* (Chicago, IL: Irwin, 1996).



Timothy R. Anderson (M'81) received the B.S. degree in electrical engineering from the University of Kentucky, Lexington, in 1977, the M.S. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1978, and the Ph.D. degree in electrical engineering from the University of Dayton, Dayton, OH, in 1990.

He is employed as Technical Advisor, Information Analysis and Exploitation Branch, Crew System Interface Division, Human Effectiveness Directorate, Air Force Research Laboratory. He has held several positions within the USAF research community since he began working for the Air Force in 1980. Prior to that time, he served as a Systems Analyst in private industry, supporting DoD R&D programs. His research interests include robust speech recognition, speaker recognition, speech synthesis, binaural hearing, and audio/visual speech processing. He has authored more than 30 journal articles, book chapters, technical reports, and published papers. He is also the coeditor of a book on binaural hearing. He holds one U.S. patent.

Edward J. Cupples (S'60–M'84) was born in Clymer, PA, on Janaury 26, 1939. He received the B.S.E.E. degree from Tri-State University, Angola, IN, in 1960.

He joined Rome Air Development Center, Griffiss Air Force Base, Rome, NY, in 1960, where he worked as a Circuit Designer and conducted studies of techniques and devices to preprocess, identify, and analyze raw data from acoustic and electromagnetic sensors to satisfy intelligence development requirements. He also engaged in programs relating to infrasonic sensing, recording and analysis, special receiver development, and the development of physical security techniques. In 1976, he joined Rome Air Development Center Audio and Speech Processing Group, where he began doing research and development in audio processing techniques. Since then, he has concentrated on the development of techniques and equipment for the processing and analysis of acoustic aircraft signals, speech enhancement, interword and intraword silence removal, and the automatic recognition of words, speakers, and languages. He is currently lead engineer in audio and speech processing research and development at Air Force Research Laboratory (formerly Rome Air Development Center).

Mr. Cupples has received numerous Air Force Scientific Achievement Awards.