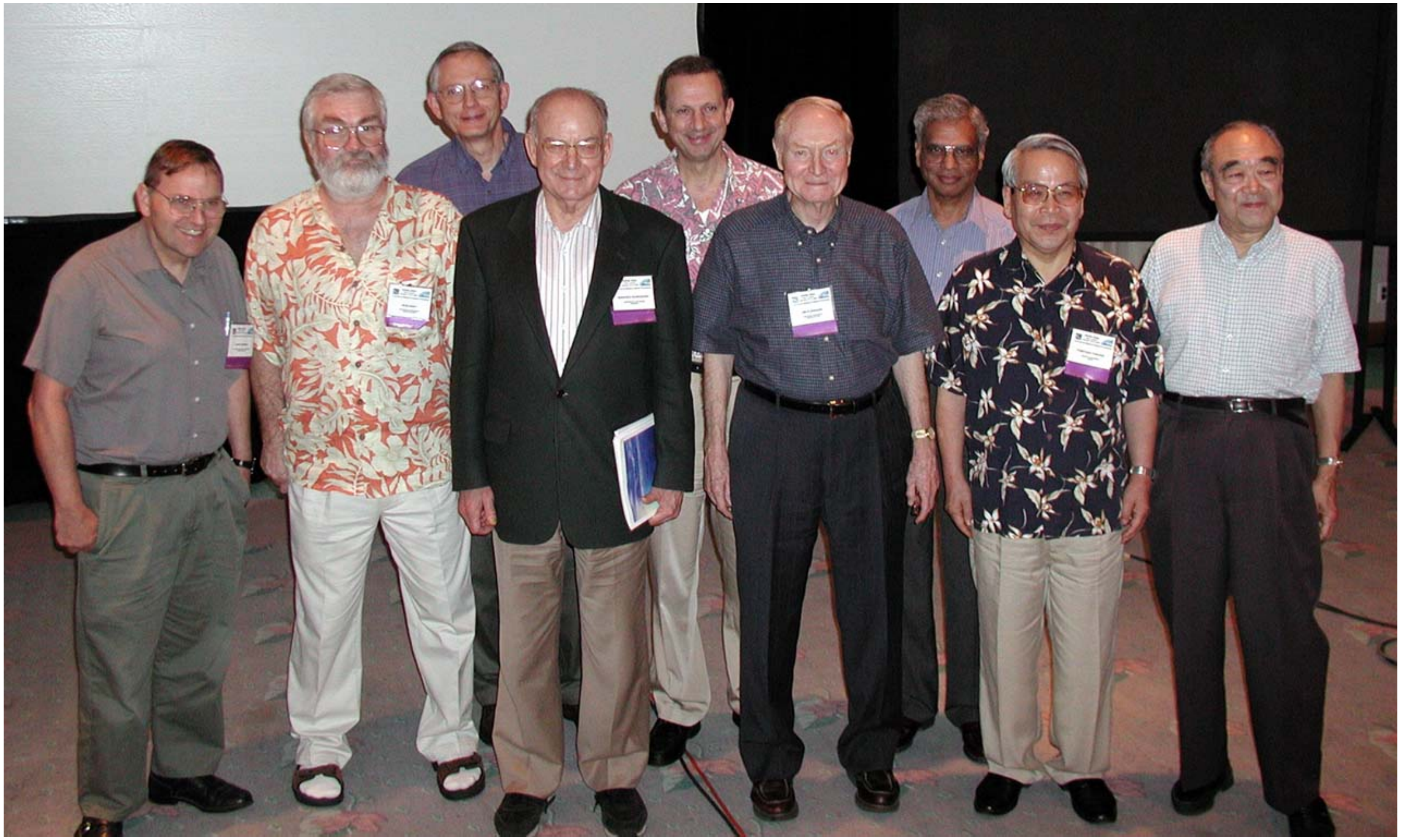
Packet speech on the Arpanet: A history of early LPC speech and its accidental impact on the Internet Protocol

Robert M. Gray
Information Systems Laboratory, Department of Electrical Engineering
Stanford, CA 94305

The author's work in speech was partially supported by the National Science Foundation. Thanks to J. D. Markel, A.H. "Steen" Gray, Jr., John Burg, Charlie Davis, Mike McCammon, Danny Cohen, Steve Casner, Richard Wiggins, Vishu Viswanathan, Jim Murphy, Cliff Weinstein, Joseph P. Campbell, Randy Cole, Rich Dean, Vint Cerf, and Bob Kahn.

Origins of this talk



Special Workshop in Maui (SWIM), 12 January 2004

Part I: Linear Prediction & Speech

Fix m. Observe a data sequence $\{X_0, X_1, \ldots, X_{m-1}\}$.

Optimal 1-step prediction ¿ What is the optimal predictor of the form $\tilde{X}_m = p(X_0, \ldots, X_{m-1})$?

Optimal 1-step linear prediction ¿ What is the optimal linear predictor of the form $\tilde{X}_m = -\sum_{l=1}^{m} a_l X_{m-l}$?

Modeling/density estimation ¿ What is the probability density function (pdf) that “best” models X^m?

Spectrum Estimation ¿ What is the “best” estimate of the power spectral density or covariance of the underyling random process?

The Application

Speech Coding ¿ How apply linear prediction to produce low bit rate speech of sufficient quality for speech understanding and speaker recognition?

Wide literature exists on all of these topics in a speech context and they are intimately related.

See, e.g., J. Makhoul's classic survey [35] and J.D. Markel and A.H. Gray's classic book [41].

Clearly problems ill-posed unless define terms like "optimal" and assume some structure.

Optimal Prediction

Random vector $X^m = (X_0, X_1, \ldots, X_{m-1})^t$

Correlation $r_{i,j} = E[X_i X_j]$, $R_n = \{r_{i,j};\ i,j = 0,1,\ldots,n-1\}$

¿ What is best $\tilde{X}_m = p(X^m)$ yielding minimum $E[(X_m - \tilde{X}_m)^2]$?

Answer: $\tilde{X}_m = E[X_m | X^m]$ MMSE $= \alpha_m = \sigma^2_{X_m|X^m}$.

If X^{m+1} Gaussian, $\Rightarrow E[X_m|X^m] = (a_{m-1}, \ldots, a_2, a_1)^t X^m$

$$(a_{m-1}, \ldots, a_2, a_1)^t = (r_{m,0}, r_{m,1}, \ldots, r_{m,m-1}) R_m^{-1}$$

$$\alpha_m = |\ R_{m+1}\ | \ / \ |\ R_m\ |$$

$\Rightarrow$ Form and performance are determined entirely by R_{m+1}!

$\Rightarrow$ optimal predictor is *linear!!*

$\Rightarrow$ optimal linear predictor = optimal predictor

Optimal Linear Prediction

Linear predictor $\tilde{X}_m = -\sum_{l=1}^{m} a_l X_{m-l}$, has MSE $\boxed{E[\epsilon_m^2] = a^t R_{m+1} a}$, where $\epsilon_m = X_m - \tilde{X}_m$ and $a \triangleq (a_0 = 1, a_1, \ldots, a_m)^t$, whether or not Gaussian.

$\Rightarrow$ optimal a for linear prediction (LP) is $\operatorname{argmin}_{a:a_0=1} a^t R_{m+1} a$ $=$ same as optimal for Gaussian! a and α_m as before.

Moral: Gaussian assumption provides short cut proofs in nonGaussian problems — no calculus and get global optimality!

Efficient inversion to find a: Cholesky decomposition $\Longrightarrow$
.
covariance method

If R_{m+1} Toeplitz, Levinson-Durbin algorithm $\Longrightarrow$
.
autocorrelation method

Other derivations: Calculus or orthogonality principle $\Rightarrow$ normal equations (Wiener-Hopf, Yule-Walker):
m linear equations in m unknowns.

¿ What if don't know R_{m+1}, but observe long sequence of actual data $X_0, X_1, \ldots, X_{n-1}$? Can estimate:

$$\hat{r}_k = \frac{1}{n-m}\sum_{l=m}^{n-1} X_l X_{l-|k|};\ \hat{R}_{m+1} = \{\hat{r}_{i-j};\ i,j = 0,1,\ldots,m\}$$

$$\overline{r}_{i,j} = \frac{1}{n-m}\sum_{l=m}^{n-1} X_{l-i} X_{l-j};\ \overline{R}_{m+1} = \{\overline{r}_{i,j};\ i,j = 0,1,\ldots,m\}$$

and "plug in." $\hat{R}_m$ Toeplitz, $\overline{R}_m$ not.

As $n \to \infty$, $\overline{R}_{m+1} \approx \hat{R}_{m+1}$

Processes and Filters

For $n = m, m+1, \ldots$ find linear least squares estimate $\tilde{X}_n = -\sum_{l=1}^{m} a_l X_{n-l}$: Previous formulation $\Rightarrow$ optimal a, MMSE α_m.

LTI filter with input X_n, response a_k: *prediction error filter* or *inverse filter* $\Leftrightarrow A(f) = \sum_{n=0}^{m} a_n e^{-i2\pi nf}$

$$X_n \rightarrow \boxed{A(f)} \xrightarrow[\text{residual, excitation}]{\epsilon_n = \sum_{k=0}^{m} a_k X_{n-k}} \boxed{1/A(f)} \rightarrow X_n = \epsilon_n - \sum_{l=1}^{m} a_l X_{n-l}$$

Limit: $m \rightarrow \infty$, the orthogonality principal $\Rightarrow$ prediction error becomes *white*!

For finite m can interpret LP as trying to make prediction error *as white as possible*.

Maximum Likelihood View

Suppose process X_n is order m Gaussian autoregressive source: W_n is iid Gaussian, variance σ^2 and

$$X_n = \begin{cases} -\sum_{k=1}^{m} a_k X_{n-k} + W_n & n = 0, 1, \ldots \\ 0 & n < 0 \end{cases}$$

Observe $X_0, \ldots, X_{n-1}$. ¿ What is maximum likelihood estimate of a, i.e., what a maximizes $f_{X^n|a}(x)$? $A_n X^n = W^n$ where

$$A_n = \begin{bmatrix} 1 & & & & & \\ a_1 & 1 & & & & 0 \\ \vdots & a_1 & 1 & & & \\ a_m & & \ddots & \ddots & & \\ & \ddots & & & & \\ 0 & & a_m & \ldots & a_1 & 1 \end{bmatrix}$$

A little algebra $\Rightarrow$

$$f_{X^n|a}(x) = \left(\frac{1}{2\sigma^2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}x^t A_n^t A_n x}$$

For large n $x^t A_n^t A_n x \approx a^t \hat{R}_{m+1} a$

Maximizing $f_{X^n|a}(x)$ over $a : a_0 = 1$ is minimizing $a^t \hat{R}_{m+1} a$

LP problem again!

ML method produces a *model* or *density estimate*: a Gaussian autoregressive process fit to data.

Maximum Entropy View

Suppose have estimate $\hat{R}_{m+1}$ of correlations to lag m of stationary random process X_n.

¿ What m-step Markov random process maximizes the Shannon differential entropy rate:

$$h(X) = \lim_{n\to\infty} \frac{1}{n} h(X^n)$$

where

$$h(X^n) = -\int f_{X^n}(x^n) \log f_{X^n}(x^n)\, dx^n\,?$$

Since assume Markov, $h(X) = h(X_m|X^m)$.

Note: No Gaussian assumption, stated as a *variational problem*.

Answer: Basic information theory $\Rightarrow$ solution is Gaussian m-th order autoregressive process with variance

$$\sigma^2_{X_m|X^m} = \frac{\mid R_{m+1} \mid}{\mid R_m \mid} = \alpha_m,$$

and power spectral density $S(f) = \alpha_m / \mid A(f) \mid^2$

LP problem again

Minimum Distortion Model Selection

Given process $\{X_n\}$ with autocorrelation R or psd S

Assume a distortion measure $d(R, R_Y)$ or $d(S, S_Y)$

Choose best model in the class $\mathcal{A}_m$ of all mth order autoregressive processes by minimum distortion (nearest neighbor) rule.

Example: Modified *Itakura-Saito distortion measure* by

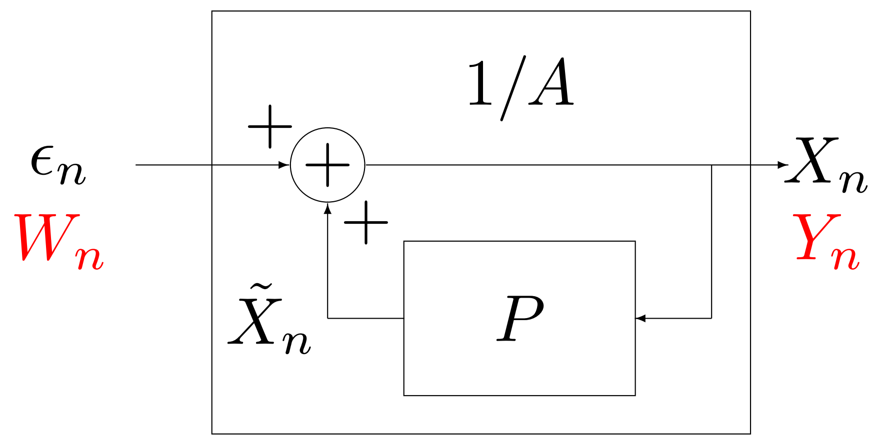
$$d(S, S_Y) = \int_{-1/2}^{1/2} \left(\frac{S(f)}{S_Y(f)} - \ln \frac{S(f)}{S_Y(f)} - 1 \right) = \frac{a^t R_{m+1} a}{\sigma_Y^2} - \ln \frac{\alpha_m}{\sigma_Y^2} - 1$$

Example of Kullback's *minimum discrimination information* for density/parameter estimation. [3, 55, 62]

Relative entropy rate between Gaussian processes (Pinsker [4])

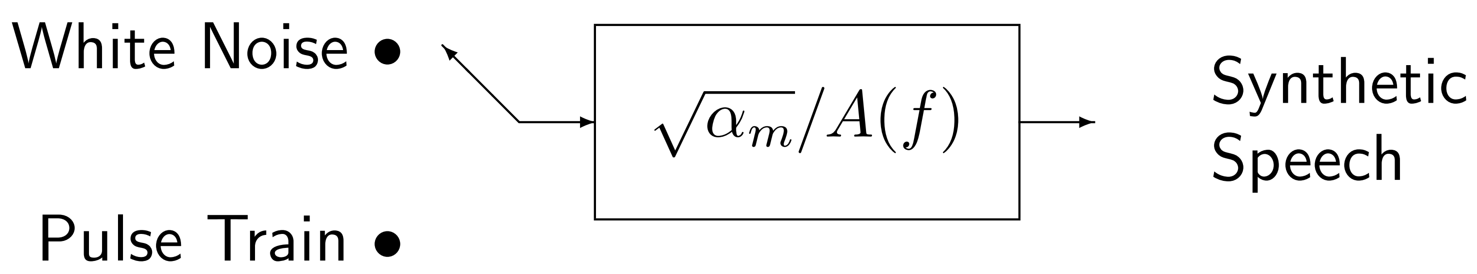
Minimized by choosing the a to minimize $a^t R_{m+1} a$ and $\sigma_Y^2 = \alpha_m$. LP problem again!

Linear Predictive Coding



$$\sigma_W^2 = \alpha_m \ , \ r_X(n) = r_Y(n), n = 0, 1, \ldots, m$$

$\Rightarrow$ LPC *model*



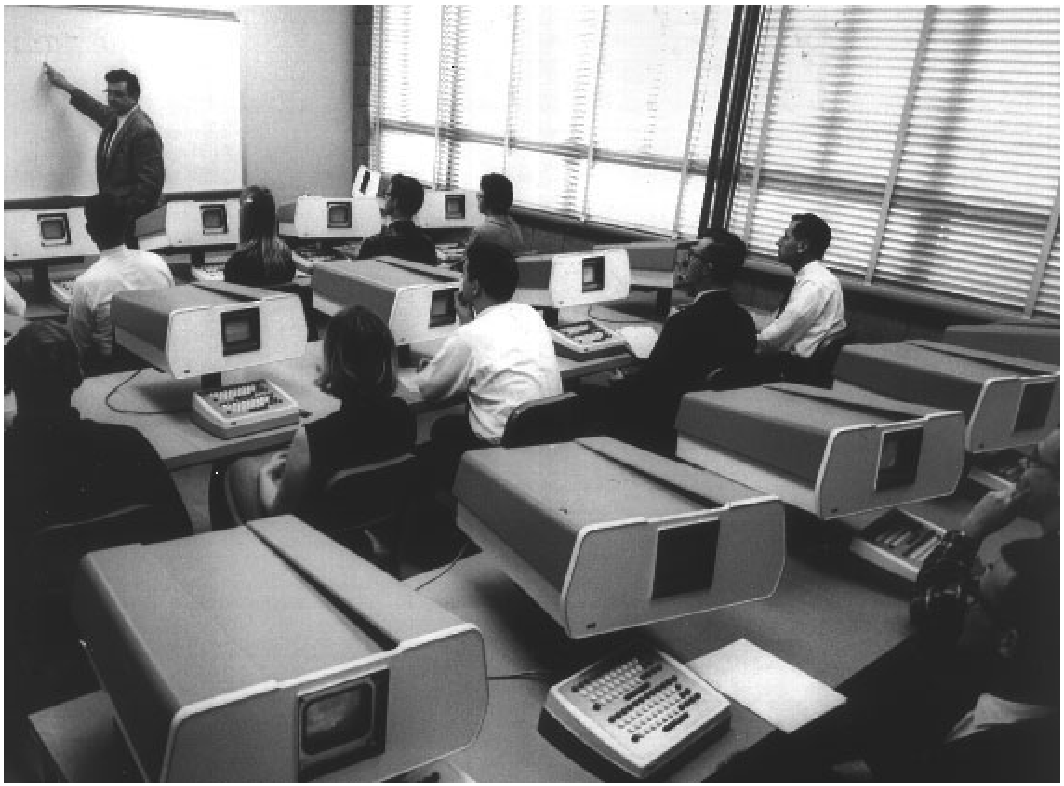
Simplistic: no voicing or pitch estimation details

LPC

Find model (α_m, A) as before. Coding occurs when the final model is selected from a discrete set, e.g., quantize separate parameters or parameter vector. Local synthesis at decoder.

Part II: History – 1966

UCSB Glen Culler introduces On-Line System (OLS or Culler-Fried system) — allows real time signal processing at individual student terminals, e.g., DFTs of real sampled speech. Culler is reknowned for building fast and effective computer systems.



1966

In December Saito and Itakura at NTT [5] describe an approach

to automatic phoneme discrimination and develop the ML approach and minimum distortion approach to speech coding: LP parameters extracted using autocorrelation method & transmitted to decoder with voicing information. Decoder synthesizes from noise or pulse train driving autoregressive filter.

See also 1968 & 1969 papers [11, 12].

From [5]:

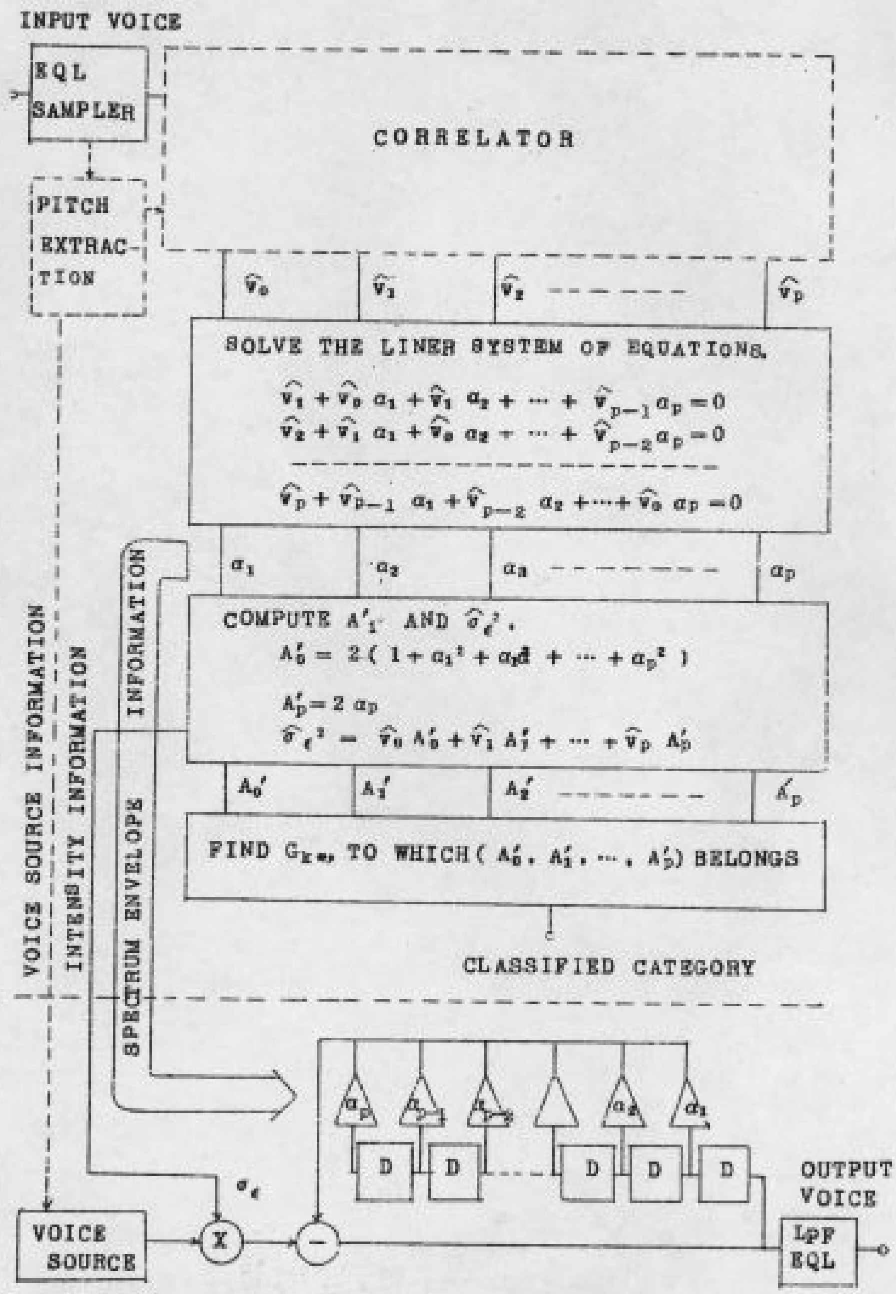


図 5. 新しいパラメータ伝送方式

−39−

1967

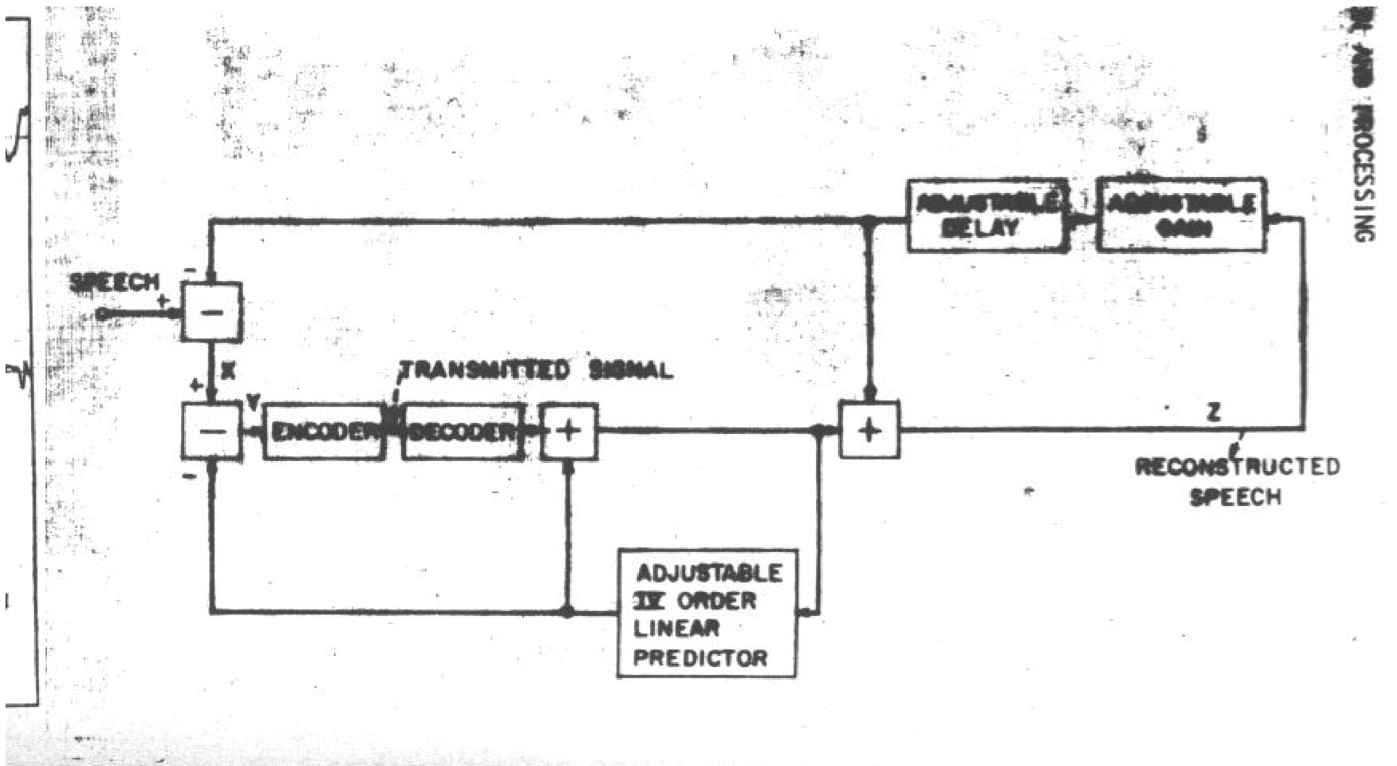
Ed Jaynes and John Burg.

October John Burg presents maximum entropy approach [9] and wins best presentation award at the meeting of the Society of Exploration Geophysists. Focus is on prediction error properties. Variational, not parametric.

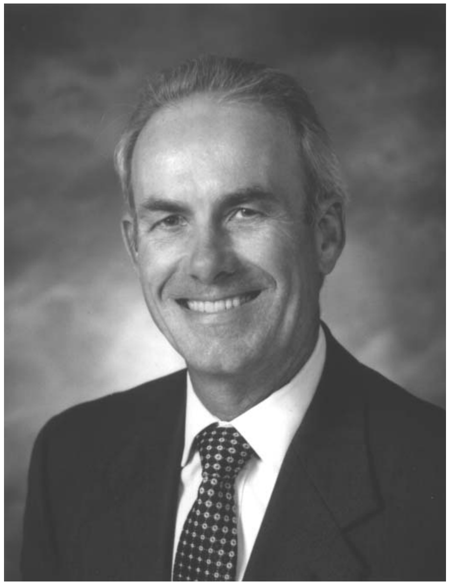
1967

November B.S. Atal and M.R. Schroeder [6]: LP coefficients used to form prediction residual, which is also coded. Adaptive predictive coding (APC), *residual excited* LPC. No explicit modeling. Elaborated in 1968 [7, 8] using covariance method.

From [6]

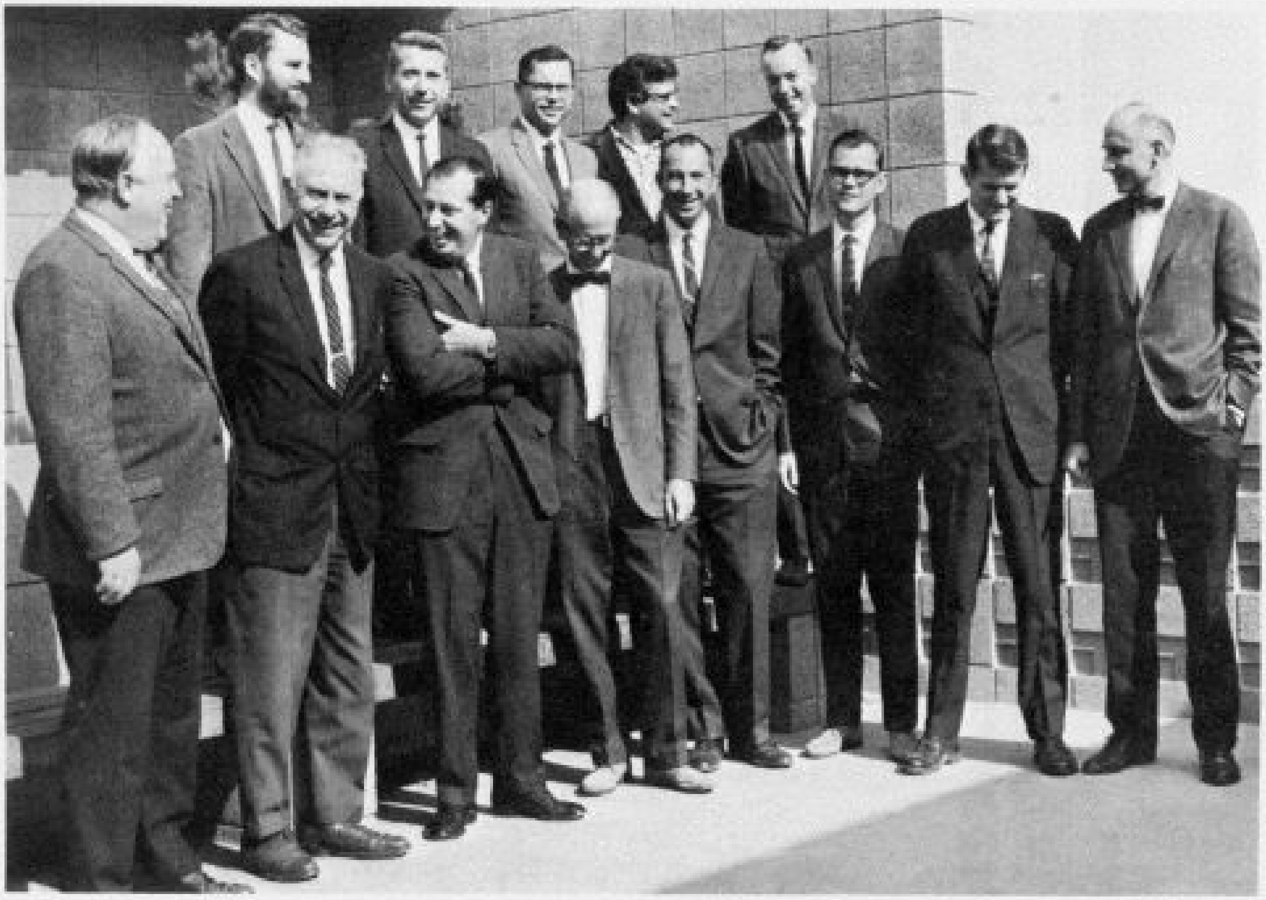


1968



John Markel drops required language course in French for PhD program at Arizona State. Moves to UCSB (Fortran is accepted there). Joins Speech Communications Research Lab (SCRL). Reads Flanagan's book and sets goal to someday write and publish a book in the same series with the same publisher.

1968



Front row: Philip Ordung, Albert Conrad, Jorge Fontana, George Matthaei, Roger Wood, Kenneth Kotzebue, John Skalnik, Glen Wade. **Second row:** Augustine Gray Jr., Glen Heidbreder, John Baldwin, Glen Culler, James Howard.

UCSB Faculty in 1968, including A.H. Gray, Jr., and Glen Culler

1968

John Burg [10] presents "A new analysis technique for time series data" at NATO Advanced Study Institute — the Burg algorithm. Finds reflection coefficients from original data using a forward-backward algorithm. Later dubbed "covariance lattice" approach in speech[46, 56].

Glen Culler contributes to Interface Message Processor (IMP) specification (with Shapiro, Kleinrock, Roberts) — the "node" of the ARPANET. BBN gets contract from ARPA to build and deploy 4 in January 1969. [72]

Culler cofounds Culler-Harrison Inc (CHI), which builds early array processors which are adopted and commercialized by FPS, will replace SPS-41 array processors.

1969

Itakura and Saito[12] introduce partial correlation (PARCOR) [1969] variation on autocorrelation method, finds partial correlation [1] coeffients. Similar to Burg algorithm, but based on classical statistical ideas and lower complexity.

May Glen Culler proposes online speech processing system aimed at real-time speech encoding based on a signal decomposition that would now be called a Gabor wavelet analysis. [13]

November B.S. Atal presents LPC speech coder at Annual Meeting of the Acoustical Society of America. [14]. Abstract published in 1970, full paper with Hanauer in 1971[16], uses covariance method.

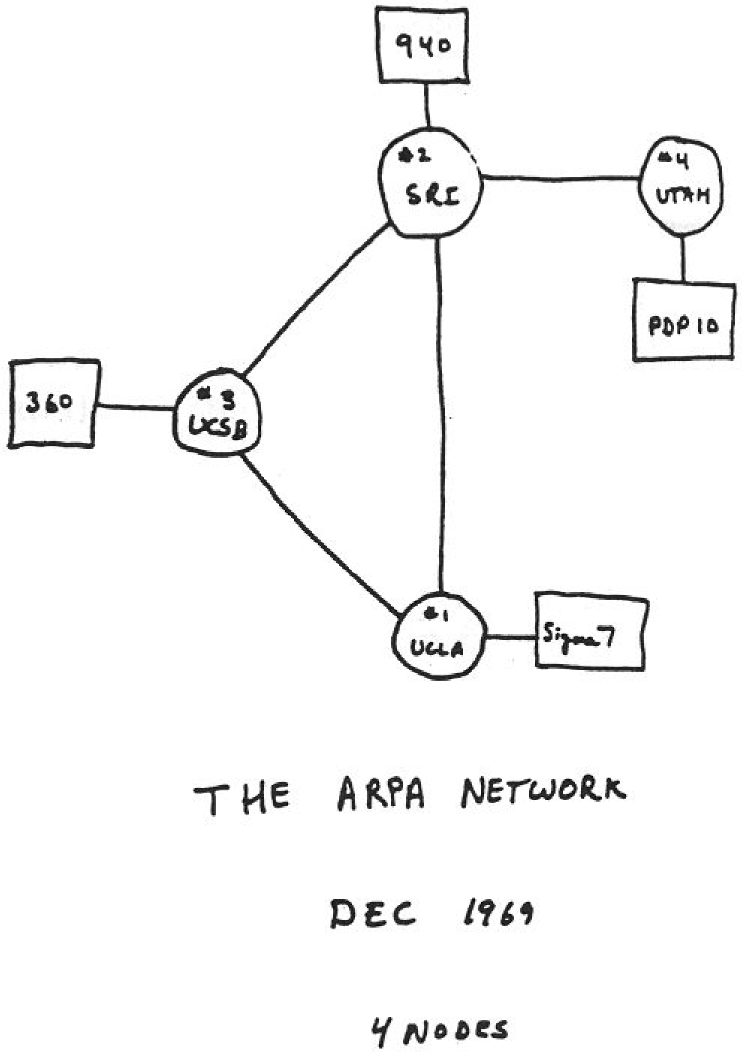


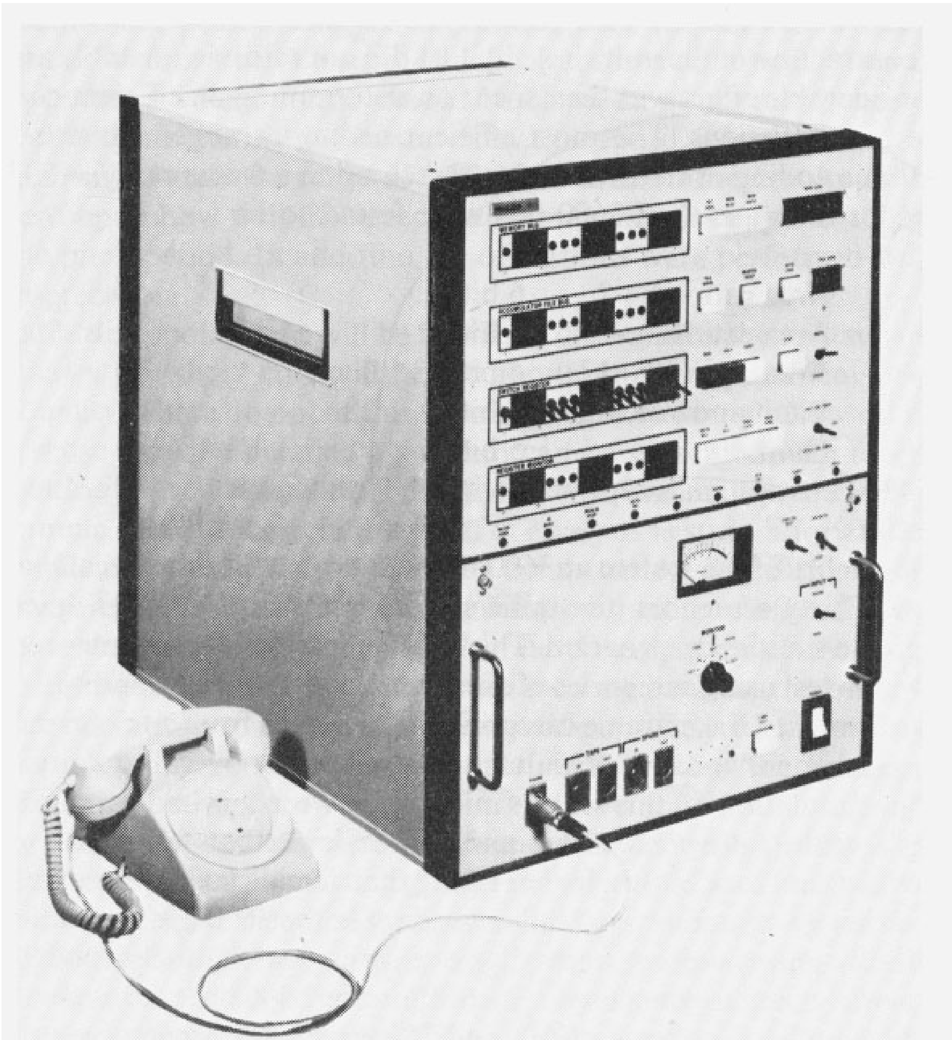
FIGURE 6.2 Drawing of 4 Node Network
(Courtesy of Alex McKenzie)

Thanks to Culler, UCSB becomes the third node (IMP) on the ARPAnet (joining #1 UCLA, #2 SRI, #4 University of Utah)

No two computers were the same (Sigma-7, SDS-940, IBM-360, and DEC-PDP10)

Drawing by Jon Postel of ISI

1971



Real time LPC using Cholesky/covariance at Philco-Ford in PA. LONGBRAKE II Final Report in 1974[33]. 16 bit fixed point LPC. Four were sold (Navy and NSA), they weighed 250 lbs @. Used PFSP signal processing computer. [41]

1972

Bob Kahn (ARPA) with Jim Forgie (LL) and Dave Walden (BBN) initiate first efforts towards packet speech on net. Simulated pieces of 64 Kbps PCM speech packets on ARPANET to understand how might eventually fit packet speech into net. Concluded major change in packet handling and serious compression would be needed.

Danny Cohen working at Harvard on realtime visual flight simulation. Bob Kahn suggests to Danny that similar ideas would work for real time speech communication over developing ARPAnet and described his project at the USC Information Sciences Institute (ISI) in Marina del Rey.

1973

Danny moves to ISI, works with Steve Casner, Randy Cole, and others and with SCRL on real time operating systems. Kahn forms Network Secure Communications (NSC) group. (Later called Network Speech Compression and Network Skiing Club because of a preference for winter meetings in Alta.) Every node on ARPAnet had different equipment and software. Focus on interface.

John Markel becomes Vice President of SCRL. Markel, Gray, and Wakita [21, 22, 23] publish SCRL reports and papers describing their implementations of Itakura's algorithms plus several of their own. Provide Fortran code for LPC and associated signal processing algorithms.

ARPA Network Information Center
Stanford Research Institute
Menlo Park, California 95025

RECEIVED
DEC 6 1973

Marcia Keeney
SRI-ARC
November 14, 1973

Network Speech Compression Note #3
NIC 19946

NETWORK SPEECH COMPRESSION GROUP MEMBERSHIP LIST

Steve F. Boll University of Utah Computer Science Dept. 3160 Merrill Engineering Building Salt Lake City, Utah 84112	SFB (801) 581-8576 UTAH-10
Dan Cohen University of Southern California Information Sciences Institute 4676 Admiralty Way Marina Del Rey, California 90291	DC (213) 822-1511 USC-ISI
Glenn J. Culler Culler-Harrison, Inc. 150-A Aero Camino Goleta, California 93017	GJC (805) 968-1813 CHI2
Robert E. Kahn Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209	REK2 (202) 694-5921 or 694-5922 ARPA-TIP
D. T. (Tom) Magill Stanford Research Institute 333 Ravenswood Avenue Menlo Park, California 94025	DTM (415) 326-6200 ext 2664 SRI
John Makhoul Bolt Beranek and Newman Inc. 50 Moulton Street Cambridge, Mass. 02138	JM (617) 491-1850 ext 234 BBN-TENEX
John D. Markel Speech Communications Research Lab, Inc. 800 Miramonte Drive Santa Barbara, California 93109	JDM (805) 965-3011 SCRL
Joseph Tierney MIT Lincoln Laboratory Lexington, Mass. 02173	JT3 (617) 862-5500 ext 277 IND

Original Members of NSC: ISI, University of Utah, BBN, MIT-LL, SRI. Soon joined by SCRL, CHI.

Others attend as well, including TI, NRL, Harris, NSA, Bell Labs

1973 Continued

Danny learns of LPC.

Remember, DSP chips did not yet exist!

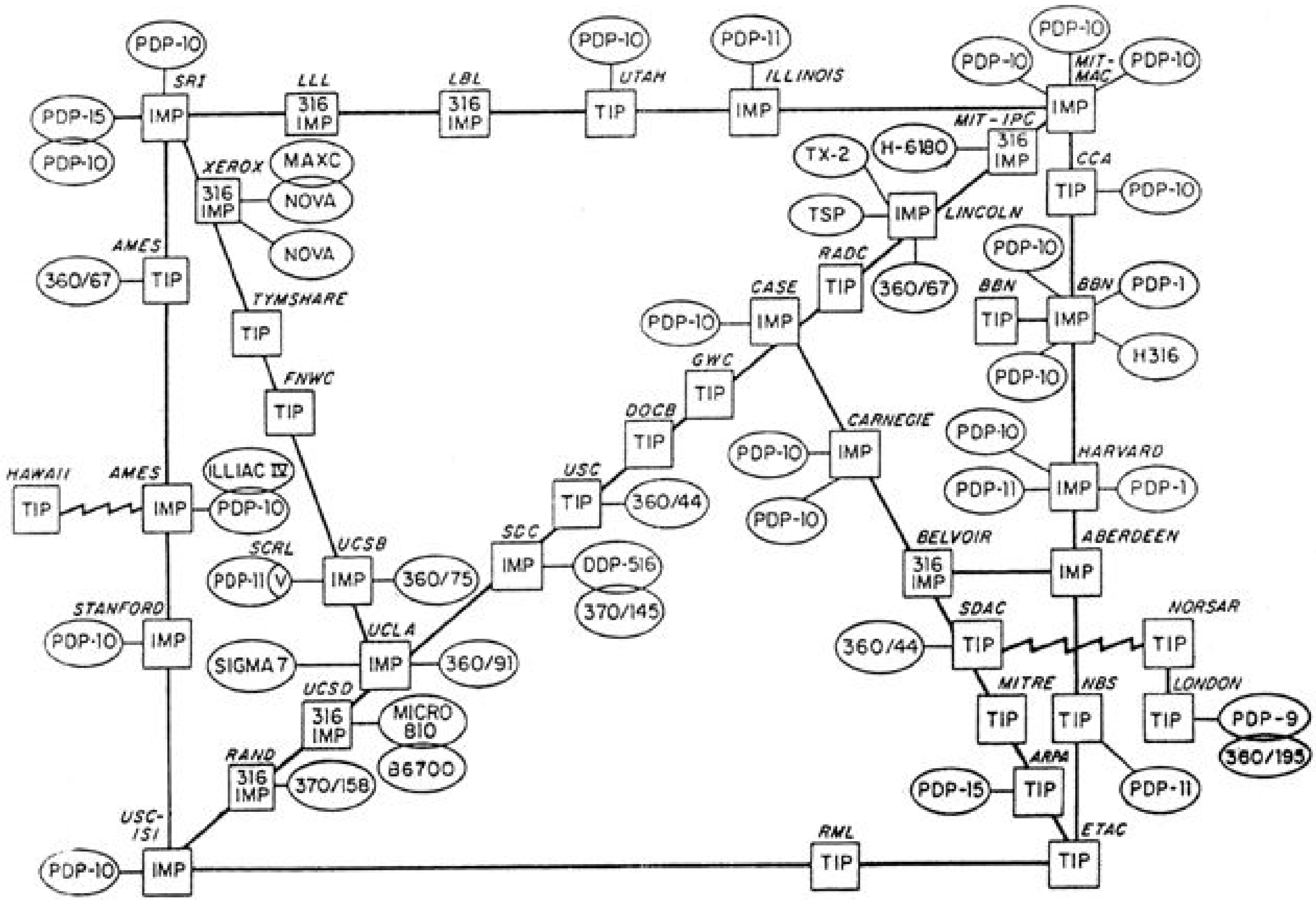
LL tradition is that first realtime 2400 bps LPC on the FDP done by Ed Hofstetter using Markel/Gray LPC formulation.

ISI adopts basic Markel/Gray software [22] as vocoder technique for network speech project. SPS-41 chosen for implementation (LL and CHI excepted) . Divided software development among ISI, BBN, LL, SRI

John Burg meets Bishnu Atal, learns of LPC.

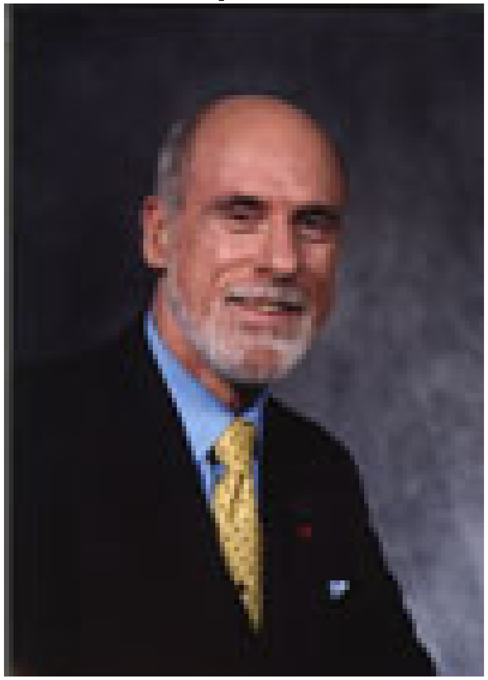
Danny Cohn raises issue of developing protocols supporting real-time applications with Kahn, who refers him to Vint Cerf.

ARPA NETWORK, LOGICAL MAP, SEPTEMBER 1973



1974

January Danny Cohen and Vint Cerf meet in Palo Alto, begin long disscussion regarding handling realtime applications vs. reliable data. Danny describes difference as the difference between milk and wine: *you have to deliver the milk quickly before it spoils even if it spills, but wine can take a LOT longer.*



TCP invented by Bob Kahn and Vint Cerf [26, 73]. Published in May.

1974 Continued

Network Voice Protocol (NVP) developed and written by Danny Cohen in coordination with others in NSC [30, 39, 43, 66]. Independent of TCP, uses only ARPANET message header.

Cohen realized that TCP was unsuitable for real time communication because of packet and reliability constraints and argued for separation of IP from the original TCP.

Implementation of TCP is a step backward for realtime applications on the ARPANET!

August NVP successfully tested using CVSD 16 Kbps, between ISI and LL. Poor quality at achievable rates.

⋆ **December 1974** First realtime two-way LPC packet speech communication. 3.5 kbs over ARPAnet between CHI and MIT-LL. [29, 34, 37, 66, 67] Uses basic M&G LPC algorithms [37, 23, 25] coupled with NVP. CHI: MP-32A signal processor + AP-90 array/arithmetic coprocessor, LL: TX2 and FDP.

Development completed on Secure Terminal Unit (STU) I. produced 1977–1979. APC using Levinson, $35K @. Speech coding at NSA led by Tom Tremain, active participant in NSC

1975

October Chaffee and Omura (UCLA) algorithm for designing a codebook of LPC models based on rate distortion theory ideas for speech coding at under 1000 bps. [31, 36]

1976

Linear Prediction of Speech by J.D. Markel and A.H. Gray, Jr. published, fulfilling Markel's goal.

Comment on M&G from Joseph P. Campbell of LL (former NSA employee): "it was considered basic reading, and code segments (translated from Fortran) from this book (e.g., STEP-UP and STEP-DOWN) are still running in coders that are in operational use today (e.g., FED-STD-1016 CELP)."

1976

Texas Instruments begins development of Speak & Spell toy: Larry Brantingham, Paul Breedlove, Richard Wiggins, and Gene Frantz.

Prior to TI, Wiggins worked on speech algorithms at MITRE in cooperation with LL and visited Itakura and Atal at Bell, NSC, Makhoul and Viswanathan at BBN, George Kang at NRL. While at TI visits Markel at SCRL and ISI in summer of 1977.

January • First LPC conference over ARPANET based on LPC and NVP successfully tested.: CHI, ISI, SRI, LL 3.5 kbps

March 1976

NVP published by Danny Cohen. Excerpt: "The Network Voice Protocol (NVP), implemented first in December 1973, and has been in use since then for local and transnet real-time voice communication over the ARPANET at the following sites:

- Information Sciences Institute, for LPC and CVSD, with a PDP-11/45 and an SPS-41.
- Lincoln Laboratory, for LPC and CVSD, with a TX2 and the Lincoln FDP, and with a PDP-11/45 and the LDVT.
- Culler-Harrison, Inc., for LPC, with the Culler-Harrison MP32A and AP-90.
- Stanford Research Institute, for LPC, with a PDP-11/40 and an SPS-41."

"The NVP's success in bridging among these different systems is due mainly to the cooperation of many people in the ARPA-NSC community, including Jim Forgie (Lincoln Laboratory), Mike McCammon (Culler-Harrison), Steve Casner (ISI) and Paul Raveling (ISI), who participated heavily in the definition of the control protocol; and John Markel (Speech Communications Research Laboratory), John Makhoul (Bolt Beranek & Newman, Inc.) and Randy Cole (ISI), who participated in the definition of the data protocol. Many other people have contributed to the NVP-based effort, in both software and hardware support."

note who is not mentioned . . .

1977

John Markel founds Signal Technology Inc. (STI). First product is Interactive Laboratory System (real time signal processing packages based on SCRL software).

WHY YOU SHOULDN'T WRITE YOUR OWN SIGNAL PROCESSING SOFTWARE.

BY A.H. "STEEN" GRAY, Jr., Ph.D.
Vice President, Signal Technology, Inc.

I've been in this business long enough to know that some things never change. The "make or buy" quandary as it applies to software is a good example. "We've got some expensive programmers; let them earn their keep." How many times have you heard that when you've suggested buying a program package that seems perfectly tailored to your application?

HOW TO ANSWER

If your line of work is signal processing, the answer should be reasonably simple. Just say, "It would take us ten years and a bundle of money to come up with a package as good as the one already available from some experts out in California." That

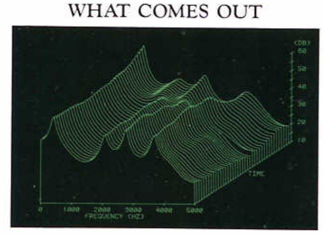
might be just a slight exaggeration, but your point would be well taken. You see, we do have the last word in interactive signal processing software. It's called ILS, and with over 200 installations worldwide, it is often referred to (and not just by us) as the "world standard."

WHAT ILS IS

ILS is a highly modular set of FORTRAN programs that make up a sophisticated interactive software system with standard file structures, documentation and ongoing support. To date, it is performing with excellent results in a wide variety of industries and technologies, including: **speech—noise and vibration—acoustics—biology—medicine—simulation—digital filtering—sonar—radar—seismic—** and some we aren't being told about.

Many of our users also find DACS, our Data Acquisition and Conversion Software, and APAS, our Array Processor Application Software, of great value in their applications.

WHAT COMES OUT



With any compatible computer system and the appropriate terminals, ILS will give you: **pattern analysis—digital filtering—signal editing—3-D displays—modeling—correlation—convolution—spectral density—signal displays—coherence—** and maybe even a picture of your Aunt Sally, if that's what you want.

THE POINT IS

The important thing is, ILS is available *now*. It has been proven in a multitude of applications around the world. And it can cost you a lot less in time and money to buy it from us rather than developing a comparable package yourself.

FREE DEMO

If you have a compatible graphics terminal and modem, we can arrange to give you an on-line demonstration of ILS. Simply call (toll free) and ask for our ILS marketing representative at (800) 235-5787. We also have a video tape that illustrates many of the features of ILS and its capabilities. We'll be happy to send it to you.

IF YOU'RE STILL WITH US

If you've read this far, you probably have some kind of interest in signal processing. I'd be delighted to send you a reprint of the series of three articles on digital filtering that I co-authored with John Markel. Write to me at the address below.

SIGNAL TECHNOLOGY, INC.
15 West De La Guerra Street, Santa Barbara, CA 93101
California (805) 963-1552 • **Outside California Call Toll Free (800) 235-5787**
TWX 910-334-3471—SIGNAL TEC SNC

Circle No. 29

1977 Continued

Development begins for STU II (LPC/APC) (1977-1980), produced 1982–1986, $13K @

April James Flanagan at Bell Labs applies for patent for "'"packet transmission of speech" four years after ARPA/NSC LL/CHI demonstration. Granted USA Patent 4,100,377 in 1978.[66, 74]

August At ISI, Cohen, Cerf, and Jon Postel discuss the need to handle real time traffic – including speech, video, and military applications. Agree to extract IP from TCP. Create user datagram protocol (UDP) for nonsequenced realtime data.

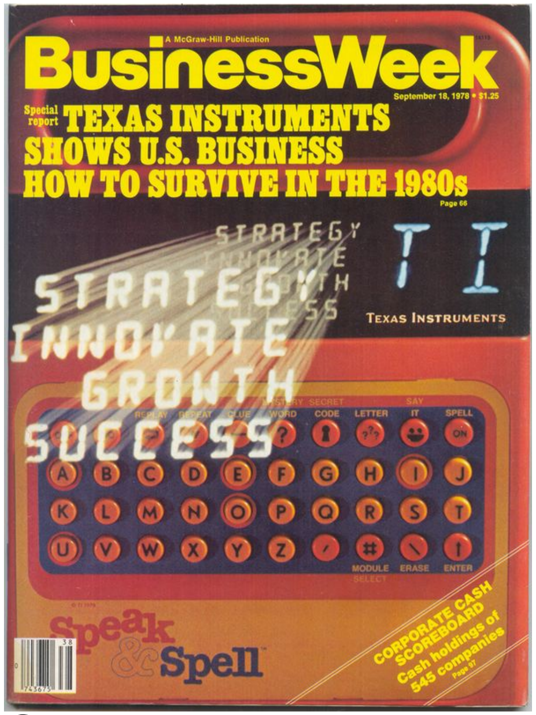
1978

January: IP officially extracted from TCP in version 3 [47].
Eventually ATM developed for similar reasons.
Finally TCP/IP suite stabilizes with version 4, still in use today.

Irony in current popular view of VoIP as novel — *IP was in fact specifically designed to handle packet speech and other realtime data!!*

April–May LPC conferencing over ARPANET using variable frame-rate (2–5 kbps) among CHI, ISI, and LL (Vishwanath et al. of BBN developed variable-rate LPC algorithm)

June Texas Instrument *Speak & Spell* toy hits the market. 1st consumer product incorporating LPC and 1st single chip speech synthesizer and early DSP chip.



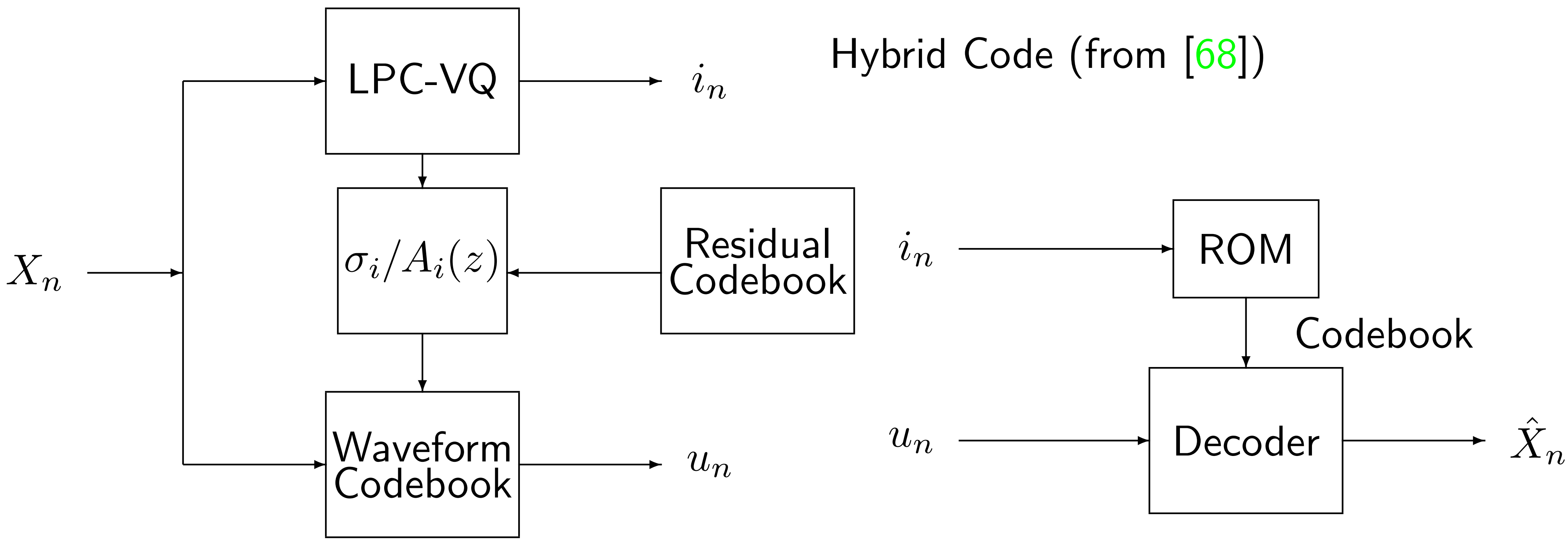
Speech synthesis from stored LPC words and phrases using TMC 0280 one-chip LPC speech synthesizer. Seminal to the development of DSP chips. Before announcement, Wiggins calls Markel, Makhoul, Atal, and Gold to acknowledge their contributions to speech and to announce the Speak & Spell. Markel asked where his royalties were — Wiggins sent him a Speak & Spell.

1979

ARPANET/Atlantic SATNET conferencing using LPC10 at 2.4kbps: ISI, LL, NDRE (Norway), UCL (London)

1981

Larry Stewart [60, 65, 68] designs trellis codes (low complexity VQ) for closed loop prediction residuals for coded LPC model. "Hybrid" code using LPC and excitation codebooks. Basic idea underlying CELP.
Simulations using legendary Altos systems at Xerox PARC.



1982

LL's single-card device using 3 NEC7720 chips, 18 sq.in. and 5.5 w shown to NSA Director. In combination with NSA's in-house development of a DSP-chip-based 2.4 kbps modem$\Rightarrow$ that secure desktop telephones were feasible, and led directly to the decision to go ahead with the STU-III development–eventually displaced all other LPC boxes.

Epilog

• Residual codebook excitation combined with perceptual weighting filters, long term prediction, and randomly generated codebook VQ to produce CELP (1985) [69] + postfiltering from Chen and Gersho (UCSB, 1987) [70] ⇒ Fed-Std-1016 CELP coder (Tremain, Campbell, Welch [71]), which included some old Markel/Gray software.

STU III: development begun in 1984, production begun 1987, \$2K each. STU III still in use, now being replaced by MELP at 2400 bps and G729 CS ACELP at 8,000 bps.

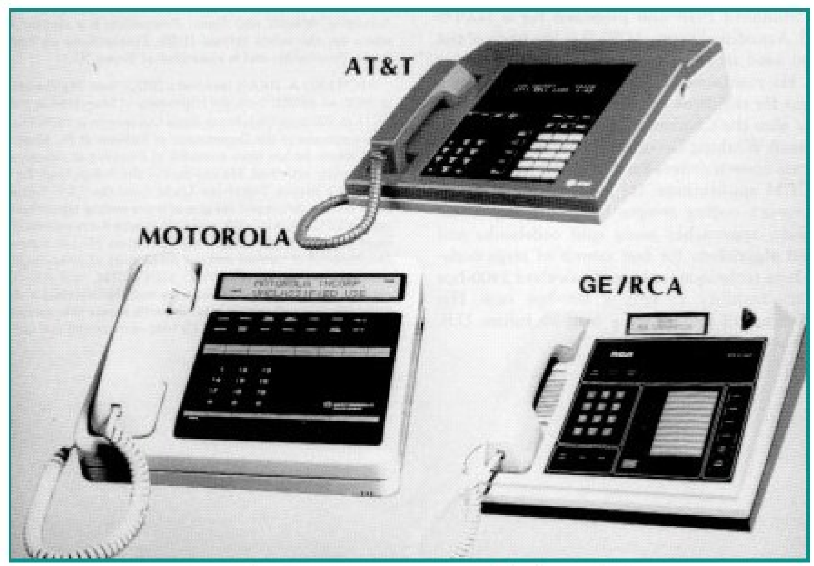


Fig. 4. The STU-III secure voice terminal family, circa 1986

Voice coding at NSA died in 2003.

•Randy Cole: “it’s hard to understate the influence that the NSC work had on networking. . . . the NSC effort was the first real exploration into packet-switched media, and we all know the effect that’s having on our lives 30 years later.”

•Barry Leiner et al. [73] “. . . some of the early work on advanced network applications, in particular packet voice in the 1970s, made clear that in some cases packet losses should not be corrected by TCP, but should be left to the application to deal with. This led to a reorganization of the original TCP into two protocols, the simple IP which provided only for addressing and forwarding of individual packets, and the separate TCP, which was concerned with service features such as flow control and recovery from lost packets.”

• In 2000 Glen Culler received the National Medal of Technology

from President Clinton for "pioneering innovations in multiple branches of computing, including early efforts in digital speech processing, invention of the first on-line system for interactive graphical mathematics computing, and pioneering work on the ARPAnet."

Culler died in May 2003.

• Vint Cerf and Bob Kahn won the 2004 ACM Turing Award for TCP/IP.

References

[1] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, John Wiley & Sons, NY, 1957.

[2] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications,* University of Calif. Press, Berkeley and Los Angeles, 1958.

[3] S. Kullback, *Information Theory and Statistics*, Dover New York, 1968. Reprint of 1959 edition published by Wiley

[4] M. S. Pinsker, "Information and information stability of random variables and processes," Holden Day, San Francisco, 1964.

[5] S. Saito and F. Itakura, "The theoretical consideration of statistically optimum methods for speech spectral density," Report No. 3107, Electrical Communication Laboratory, NTT, Tokyo, December 1966.

[6] B.S. Atal and M. R. Schroeder, "Predictive coding of speech signals," *Proc. 1967 AFCRL/IEEE Conference on Speech Communication and Processing*, pp. 360–361, Cambridge, Mass, 1967.

[7] B.S. Atal and M. R. Schroeder, "Predictive coding of speech signals," *Rep. 6th Int. Congr. Acoust.*, Y. Konasi, Ed., Tokyo Japan, Rep.C-5-5, August 1968,

[8] B.S. Atal and M. R. Schroeder, "Predictive coding of speech signals," *WESCON Tech. Papers*, Paper 8/2, 1968.

[9] John Parker Burg, "Maximum entropy spectral analysis," presented at the 37th Meeting of the Society of Exploration Geophysisists, Oklahoma City, Oklahoma, October 1967.

[10] John Parker Burg, "A new analysis technique for time series data," presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics," Enschede, The Netherlands, Aug. 1968, reprinted in *Modern Spectrum Analysis*, D. G. Childers, ed., IEEE Press, New York, 1978.

[11] F. Itakura and S. Saito, "Analysis synthesis telephony based upon the maximum likelihood method," Reports of 6th Int. Cong. Acoust., ed. by Y. Kohasi, Tokyo, C-5-5, C17–20, 1968.

[12] F. Itakura and S. Saito, "Analysis synthesis telephony based on the partial autocorrelation coefficient," Acoust. Soc. of Japan Meeting, 1969.

[13] Glen J. Culler, "An Attack on the Problems of Speech Analysis and Synthesis with the Power of an On-Line System," *Proceedings of the 1st International Joint Conference on Artificial Intelligence*, Donald E. Walker, Lewis M. Norton (Eds.): Washington, DC, May 1969. pp. 41–48.

[14] B.S. Atal, "Speech analysis and synthesis by linear prediction of the speech wave," presented at the 78th Meeting of the Acoustical Society of America, San Diego, November 1969. Abstract in *J. Acoust. Soc. Am.*, Vol 47, p. 65, 1970.

[15] "Adaptive predictive coding of speech signals," *Bell Sys. Tech. J.*, Vol. 49, No. 8, pp. 1973–1986, October 1970.

[16] B. S. Atal and S.J. Hanauer "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoustic Society of America*, Vol. 50, pp. 637–655, August 1971.

[17] F. Itakura and S. Saito, "Digital filtering techniques for speech analysis and synthesis," *Conf. Rec. 7th Int Congr Acoustics*, Paper 25C1, 1971.

[18] F. Itakura and S. Saito, "On the optimum quantization of feature parameters in the parcor speech synthesizer," *Conference Record, 1972 International Conference on Speech Communication and Processing*, Boston, MA, pp. 434–437, April 1972.

[19] J.D. Markel, "Digital inverse filtering–a new tool for formant trajectory estimation," *IEEE Trans. on Audio and Electro Acoustics*, pp. 129–13?, June 1972. (autocorrelation method)

[20] H. Wakita, "Estimation of the vocal tract shape by optimal inverse filtering and acoustic/articulatory conversion methods," *SCRL Monograph No. 9*, SCRL, Santa Barbara, CA 1972.

[21] J.D. Markel, A.H. Gray, and H. Wakita, *Linear Prediction of Speech — Theory and Practice*, SCRL Monograph No. 10, Speech Communications Research Laboratory, Santa Barbara, California, 1973.

[22] "Documentation for SCRL Linear Prediction Analysis/Synthesis Programs," Speech Communications Research Labs, Inc." Nov. 1973.

[23] J.D. Markel and A.H. Gray, Jr., "On autocorrelation equations as applied to speech analysis," *IEEE Trans. on Audio and Electroacoustics*, Vol. 21, pp. 69–79, April 1973.

[24] D.T. Magill, "Adaptive speech compression for packet communication systems," *Conf. Rec. 1973 IEEE Telecom. Conf. Proc*, 29D-129D-5. DELCO algorithm, LPC with removal of silence and sending only changes.

[25] J.D. Markel and A.H. Gray, Jr., "A linear prediction vocoder simulation based upon the autocorrelation method," *IEEE Trans on Acoustics, Speech, and Signal Processing*, Vol. 22, pp. 124–134, April 1974.

[26] Vinton Cerf, Robert Kahn, "A protocol for packet network internetworking," *IEEE Trans. Commun.*, Vol. 22, 627–641, May 1974.

[27] Draft Summary of First CHI–LL LPC Attempt Mike McCammon 11/26/74

[28] Draft Experiment #2 CHI–LL LPC Voice Contact Mike McCammon 11/26/74

[29] Report on Third CHI–LL LPC Voice Experiment 26 November 10:00–1:00 AM PST Mike McCammon

[30] ISI/SR-74-2 Annual Technical report May 1973-June 1974

[31] D.L. Chaffee and J.K. Omura, "A very low rate voice compression system," *Abstracts of Papers, International Symposium on Information Theory*, p. 69, October, 1974.

[32] NSC Note 47 Two NSC Group Proposals and a Status Report J.D. Markel SCRL 11 November 1974

[33] J. Welch, "LONGBRAKE II Final Report," Contract No. DAAB03-74-C-0098, Philco-Ford Corporation, Willow Grove, Pennsylvania, 1974.

[34] ISI/SR-75-3 Annual Technical report May 1974-June 1975

[35] John Makhoul, "Linear prediction: a tutorial review," *Proceedings of the IEEE*, Vol. 63, No. 4, April 1975.

[36] D.L. Chaffee, *Applications of rate distortion theory to the bandwidth compression of speech signals*, PhD Dissertation, UCLA.

[37] Glen J. Culler, Michael McCammon, and J.F. McGill, "Real time implementation of an LPC algorithm," *Culler/Harrison Inc. Quarterly Technical Report on Speech Signal Processing Research at CHI, during Nov. 1974–April 1975*, May 1975.

[38] Cheong Un and D. Thomas Magill, "The residual-excited linear prediction vocoder with transmission rate below 9.6kbs, *IEEE Trans on Communications*, Vol 23, pp. 1466-1474.

[39] D. Cohen "Specifications for the Network Voice Protocol," USC/Information Sciences Institute, ISI/RR-75-39, March 1976. Also NSC Note 68, January 1976.

[40] ISI/SR-76-6 Annual Technical report July 1975 – June 1976

[41] J.D. Markel and A.H. Gray, Jr, *Linear Prediction of Speech*, Springer-Verlag, 1976.

[42] A.H. Gray and J.D. Markel, "Distance Measures for Speech Processing," *IEEE Trans. on ASSP*, Vol. 24, no. 5, Oct. 1976.

[43] D. Cohen, "RFC0741: Specifications for the Network Voice Protocol," 22 Nov 1977.

[44] A. H. Gray, Jr., R. M. Gray and J. D. Markel, "Comparison of optimal quantizations of speech reflection coefficients," *IEEE Trans. on Acous., Speech & Signal Process.*, Vol. ASSP–25, pp. 9–23, Feb. 1977.

[45] Gold, Bernard (invited paper), "Digital Speech Networks", Proc. IEEE Vol. 65, No. 12, Dec. 1977. (Discusses ARPA voice project.)

[46] J. Makhoul, "Stable and lattice methods for linear prediction," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 25, pp. 423–428, October 1977.

[47] "SPECIFICATION OF INTERNETWORK TRANSMISSION CONTROL PROGRAM TCP Version 3," Vinton G. Cerf, Advanced Research Projects Agency, and Jonathan B. Postel, Information Sciences Institute January 1978.
Internet Experiment Note 21, January 1978.

[48] Wiggins, R. and L. Brantingham, "Three-chip system synthesizes human speech," Electronics, Aug 31, 1978, 109-116. Uses TMC 0280, LPC-10, 600-2400 bps.

[49] Y. Matsuyama, "Speech compression systems using a set of inverse filters," PhD Dissertation, Stanford University, August 1978.

[50] R.M. Gray, A.Buzo, Y. Matsuyama, A.H. Gray, Jr. and J.D. Markel, "Source coding and speech compression," *Proceedings of the 1978 Int'l. Telemetering Conf., 24*, Los Angeles, CA, pp. 871–878, Nov. 1978.

[51] A. Buzo, R.M. Gray, A.H. Gray, Jr., and J.D. Markel, "Optimal Quantizations of Coefficient Vectors in LPC Speech," *1978 Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan*, Honolulu, HI, Dec. 1978.

[52] A. Buzo, A.H. Gray, Jr., R.M. Gray, and J.D. Markel, "A Two–Step Speech Compression System with Vector Quantizing," *Proceedings of the 1979 Int'l. Conf. on Acoustics, Speech and Signal Processing*, pp. 52–55, Wash. DC, 1979.

[53] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. on Comm.*, Vol. COM–28, pp. 84–95, Jan. 1980.

[54] J.-P. Adoul, J.L. Debray, and D. Dalle, "Spectral distance measure applied to the design of DPCM coders with L predictors," *Conf. Record 1980 IEEE ICASSP*, Denver, CO, pp. 512–515, April 1980.

[55] R. M. Gray, A. Buzo, A. H. Gray, Jr., and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Trans. Acous., Speech and Sign. Proc.*, Vol. ASSP–28, pp. 367–376, Aug. 1980.

[56] A.H. Gray, Jr., and D.Y. Wong, "The Burg algorithm for LPC Speech analysis/synthesis," *IEEE Trans on Acoustics, Speech, and Signal Processing,*"Vol. 28, No. 6, December 1980, pp. 609–615.

[57] A. Buzo, A.H. Gray, Jr., R.M. Gray, and J.D. Markel, "Speech coding based upon vector quantization," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP–28, pp. 562–574, October 1980.

[58] Cohen, D., "A voice message system," in R. P. Uhlig (ed.), Computer Message Systems, pp. 17-27, North-Holland, 1981. (Discusses ARPA voice project.)

[59] Y. Matsuyama and R.M. Gray, "Universal tree coding for speech," *IEEE Trans. on Information theory*, Vol. 27, No. 1, pp. 31–40, January 1981.

[60] L. C. Stewart, *Trellis Data Compression*, Ph.D. Dissertation, Department of Electrical Engineering, Stanford University, June 1981.

[61] J. van Campenhout and T. Cover, "Maximum entropy and conditional probability," *IEEE Transactions on Information Theory*, Volume: 27 , No 4 , July pp 483 – 489, 1981.

[62] R.M. Gray, A.H. Gray, Jr., G. Rebolledo, and J.E. Shore, "Rate distortion speech coding with a minimum

discrimination information distortion measure," *IEEE Transactions on Information Theory*, vol. IT–27, no. 6, pp. 708–721, Nov. 1981.

[63] "An 800 b/s vector quantization LPC vocoder," D.Y. Wong, J.-H. Juang, and A.H. Gray, Jr., *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 30, pp. 770–780, 1982.

[64] D.Y. Wong and J.-H. Juang *Vector/matrix quantization for narrow-bandwidth digital speech compression*, Final Report summarizing contract No. F30602-81-C-0054, Signal Technology, Inc., July 1982.

[65] L.C. Stewart, R.M. Gray, and Y. Linde, "The design of trellis waveform coders," *IEEE Transactions on Communications*, Vol. COM–30, pp.702–710, April 1982.

[66] Excerpts from "Packet Speech Program Review Meeting" 3 June 1982 Sponsored by DARPA Hosted by MIT LL "Network voice protocols," Danny Cohen ISI, pp. 40–59 'Packetized Speech Overview" pp. 17–23

[67] C.J. Weinstein and J.W. Forgie, "Experience with speech communication in packet networks," *IEEE Journal on Selected Areas in Communications*, Vol. 1, No. 6, December 1983.

[68] R.M. Gray, "Vector Quantization," *IEEE ASSP Magazine*, Vol. 1, pp. 4–29, April 1984.

[69] M.R. Schroeder and B. Atal, Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates, Proc. ICASSP-85, p. 937, Tampa, Apr. 1985.

[70] Chen and A. Gersho, "Real-Time Vector APC Speech Coding at 4800 bps with Adaptive Postfiltering," *Proc. ICASSP '87*, pp. 2185–2188.

[71] J. P. Campbell, T. E. Tremain, and V. C. Welch, "The Federal Standard 1016 4800 bps CELP Voice Coder," *Digital Signal Processing*, Vol. 1, no. 3 (1991): 145 – 155.

[72] J. Salus, "Casting the Net: From ARPANet to Internet and Beyond," Addison-Wesley, Reading, Mass., 1995.

[73] B.M. Leiner, V.S. Cerf, D.D. Clark, R.E. Kahn, L. Kleinrock, D.C. Lynch, J. Postel, L. Kleinrock, D.C. Lynch, J. Postel, L. C. Roberts, S.S. Wolff, "The past and future history of the Internet," *Communications of the ACM*. Vol. 40, No. 2, 102–108, February 1997. See also `http://www.isoc.org/internet-history/`

[74] Danny Cohen, "Realtime Networking and Packet Voice," SIGCOM'99

[75] R. M. Gray, "Digital Speech and the Internet Protocol: The 1974 Origins of VoIP," *IEEE Signal Processing Magazine*, Vol. 22, July 2005, pp. 87–90.