Intelligibility and Space-based voice with relaxed delay constraints

Sam Nguyen, Clayton Okino, and Michael Cheng

Jet Propulsion Laboratory

Presented at

IEEE Aerospace Conference

Big Sky, Montana

5 March 2008

Outline

- Background: Space communications considerations
- Luby-Transform (LT) Codes
- Metrics used in testing & experimental setup
- Results
- Intelligibility Overview
- Results
- Conclusions
- Future directions

Space Communications Characteristics

- End-to-end latency is significant relative to the terrestrial environment
  - E.g. ~1.3 sec one-way propagation delay Moon-Earth
- Wireless communications channels are potentially noisy resulting in bit errors and/or dropped packets
- Automatic retransmission query (ARQ) techniques rely on a return channel (feedback) which may undesirable and impose to high a constraint versus a sufficient simplex channel need
  - Operation over simplex channel
  - Tolerate errors, or exploit error concealment techniques

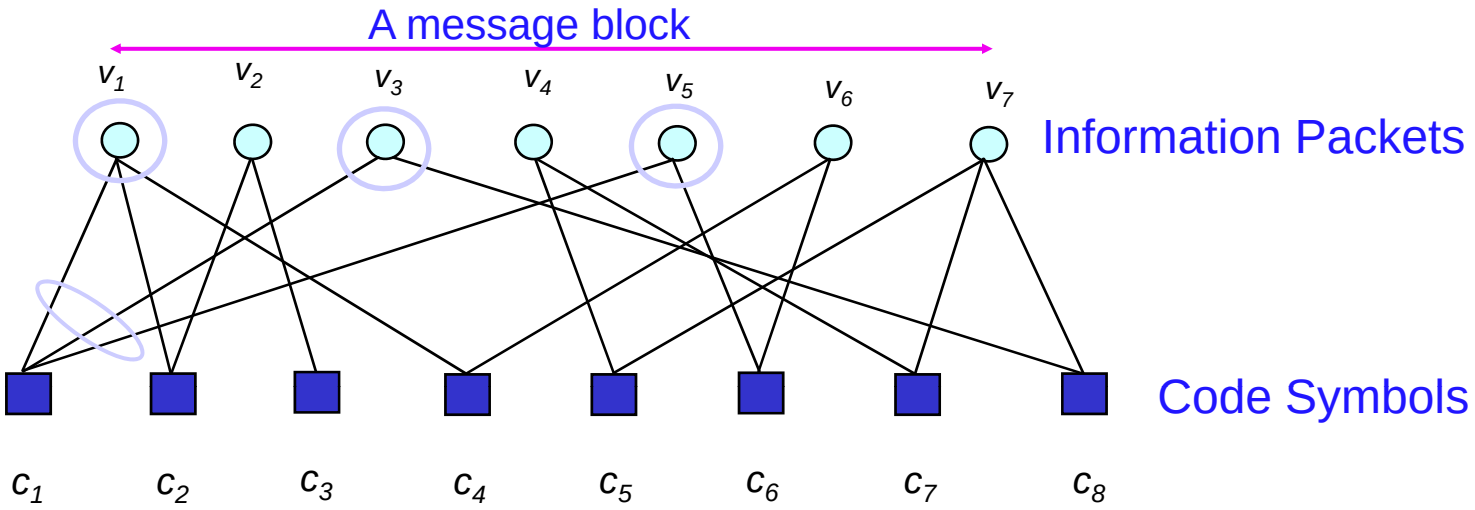
Terrestrial Networks

- **Lower Latency**
- **Lower BER**
- **Can Request Resend on Error**

Space Networks

- **Higher Latency**
- **Higher BER**
- **Require Anticipatory Error Recovery**

Encoder for LT codes



For each code symbol:

1. Randomly select the number of information packets to be XORed according to the robust soliton distribution. Example: 3 bits for symbol c_1.
2. Randomly select the positions of the information packets to be XORed according to a uniform distribution. Example: positions 1, 3, 5, for symbol c_1.
3. XOR the selected bits to generate the code symbol. Example: $c_1=v_1+v_3+v_5$.

Decoders for LT codes

Algebraic decoder:

Each code symbol establishes a constraint with the information packets in a message block. So a collection of code symbols establishes a system of linear equations. Solution to this system of equations is the original information packets.

$$\underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & \cdots & 0 \\ 1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & & & \cdots & \vdots \end{bmatrix}}_{G} \underbrace{\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix}}_{\underline{v}} = \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \\ \vdots \end{bmatrix}}_{\underline{c}}$$

1. Collect code symbols $\underline{\boldsymbol{c}}$ until $\boldsymbol{G}$ is full rank.
2. Recover $\underline{\boldsymbol{v}}$ by computing $\boldsymbol{G}^{-1}\underline{\boldsymbol{c}}$.

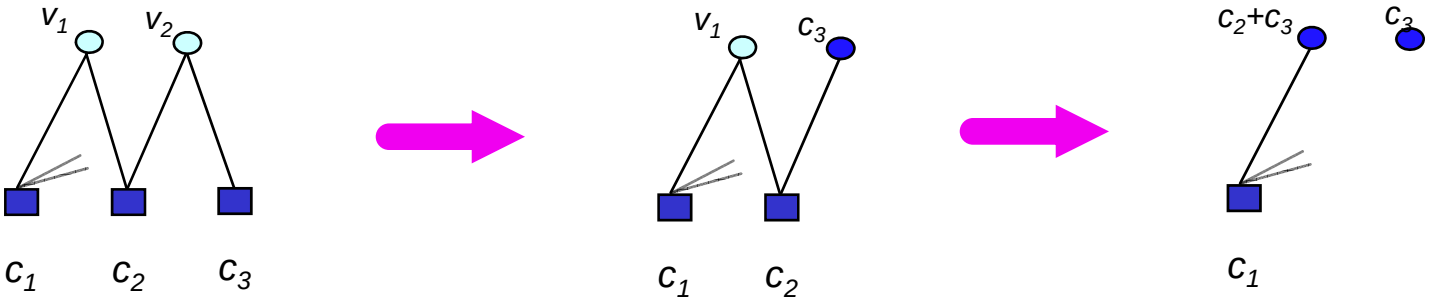
Advantage: low average over head.

Disadvantage: inverting a matrix is of complexity $O(k^3)$.

Decoders for LT codes (cont.)

Belief Propagation (BP) decoder:

1. Find a code symbol c_i that is connected to only one information packet v_j. (If there is no such code symbol, the decoder halts and declares a decoder failure).
2. Set $v_j=c_i$.
3. Add v_j to all code symbols c_i's that are connected to v_j.
4. Remove all edges connected to the information packet v_j.
5. Repeat steps 1-4 until all information packets are recovered.

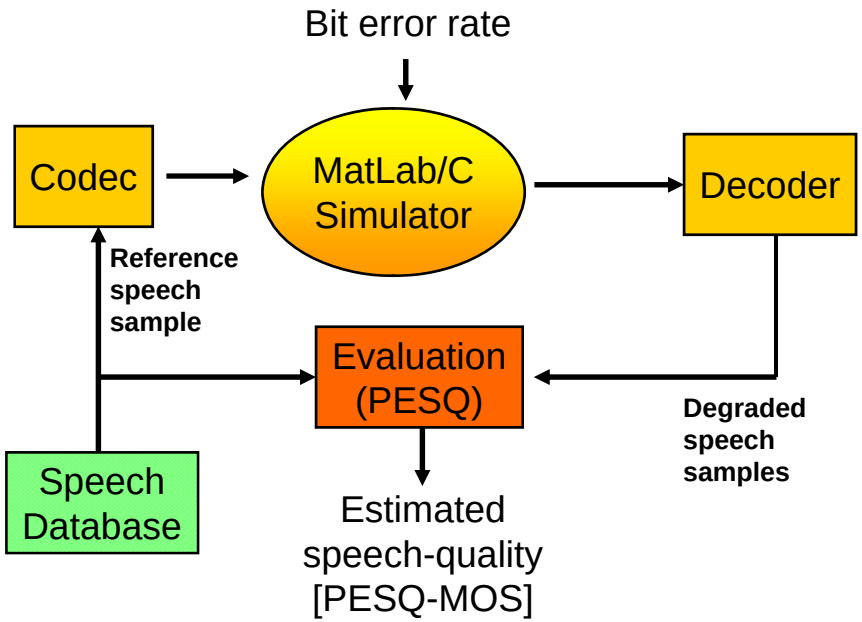


Advantage: decoding complexity is ~***O(klogk).***

Disadvantage: average overhead is higher than the algebraic decoder.

Metrics Used & Experimental Set Up

- **Speech Quality**
  - **Perceptual Evaluation of Speech Quality (PESQ) algorithm provides an objective measure of pf speech quality.**
  - **This is as opposed to the Mean Opinion Score (MOS) subjective approach.**
  - **The basic simulation modeling approach is used from Florian Hammer and is shown below**



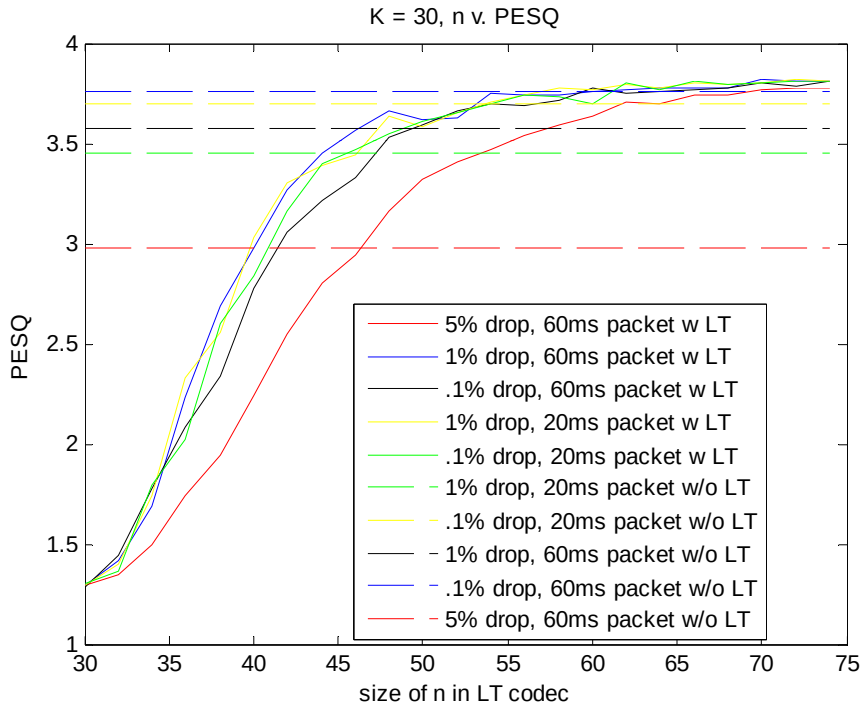
Codec

- **Codec analysis did not encompass all possible candidates and work focused on one codec as a initial assessment**
  - **Selected codec has good PESQ performance for bandwidth efficiency but is not necessarily the optimal choice**
  - **As described in [kataoka] G.729 codec is an 8 kbps conjugate structure code excited linear prediction algorithm (CS-CELP)**
    - **Operates on 10 ms blocks of encoded speech**
    - **Utilizes linear predictive coding analysis**
    - **Utilizes codebooks for the set of possible sequences**
    - **Conjugate relationship between two codebooks used for the random excitation vector**
      - **Similar relationship for the gain vector**

[kataoka] A. Kataoka, T. Moriya, “An 8 kb/s Conjugate Structure CELP (CS-CELP) Speech Coders”, IEEE Transactions on Speech and Audio Processing , Vol. 4, No. 6, November 1996.

Results

- **G.729 CODEC PESQ performance degrades at various size of LT codes to number of 10ms frame per packet**

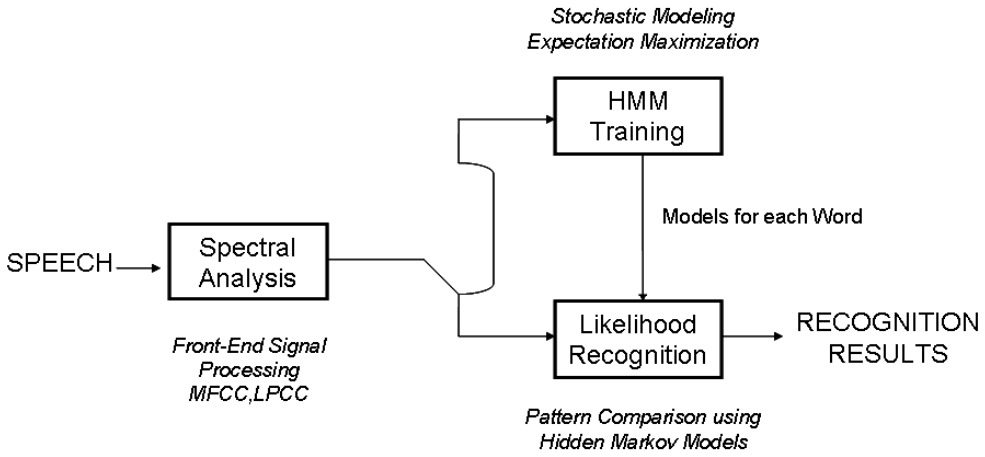


Intelligibility Overview

- **Dynamic Rhyme Test**

Voicing	Nasality	Sustenation
Veal-Feel	Meat-Beat	Vee-Bee
Bean-Peen	Need-Deed	Sheet-Cheat
Gin-Chin	Mitt-Bit	Vill-Bill
Dint-Tint	Nip-Dip	Thick-Tick
Zoo-Sue	Moot-Boot	Foo-Pooh

- **Speech Recognition**



Results

- **Dynamic Rhyme Test**

Speaker	DRT Score	Standard Error
RH	96.9	.74
JE	93.9	.72
CH	96.4	.96
VW	95.6	.55
KS	98.0	.69
MP	97.5	.39

- **Speech Recognition**

Speaker	#correctly identified	#wrongly Identified	% of words correctly identified
RH	172	20	89.58
JE	161	31	83.85
CH	167	25	86.98
VW	141	51	73.44
KS	156	36	81.25
MP	150	42	78.13

Conclusions

- Utilizing LT codes as a means of reducing packet erasures due to corrupted packets on an RF link can result in higher voice quality
  - E.g. Tolerating 720 ms of delay can result in error-free G.729 performance for a 5% packet drop rate channel
- ASR as a means of obtaining a metric related to DRT is a promising area for further work
- PESQ-MOS measure was used to analyze voice degradation over space links tested for LT codec size and number of 10ms per packet

Future Directions

- **Extensions utilizing LT codes to improve the packet erasure performance and combining the use of ASR could provide for a solid means of identifying the benefit in terms of intelligibility of voice communications in space-based networks**