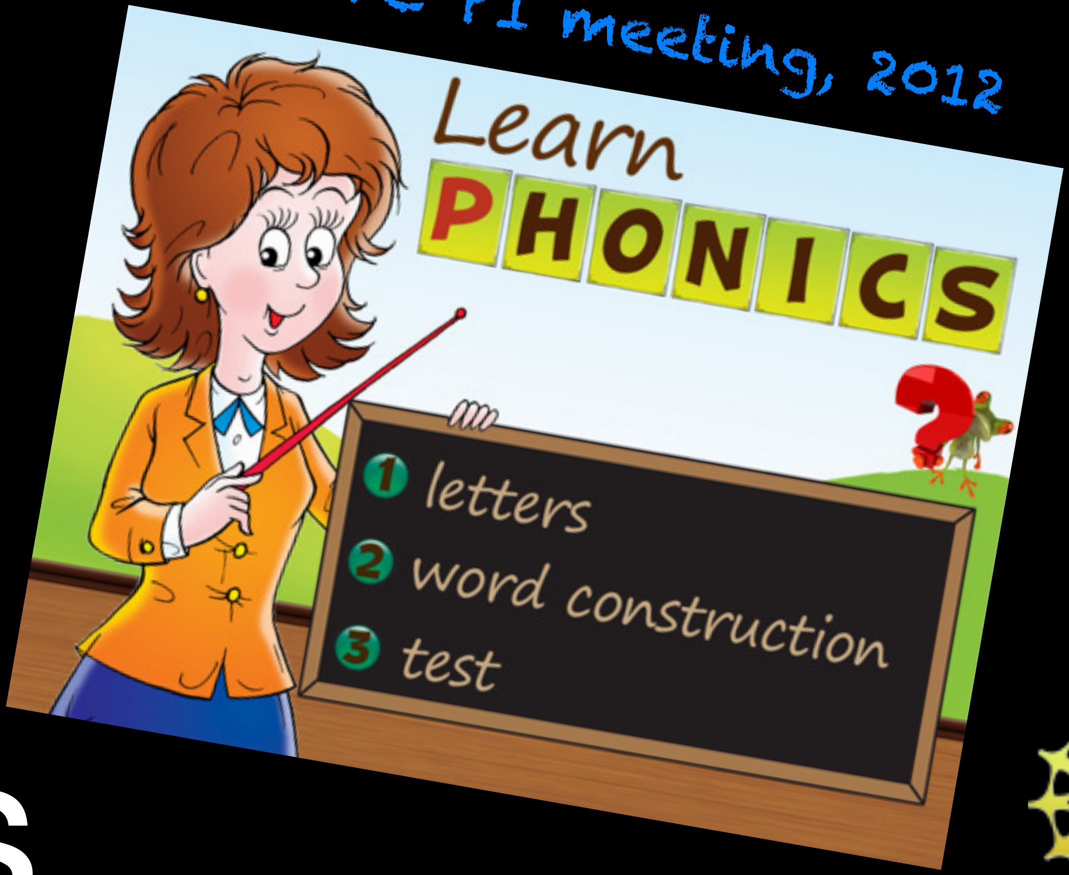
SaTC PI meeting, 2012



hʊkt ɔn fɒnɪks

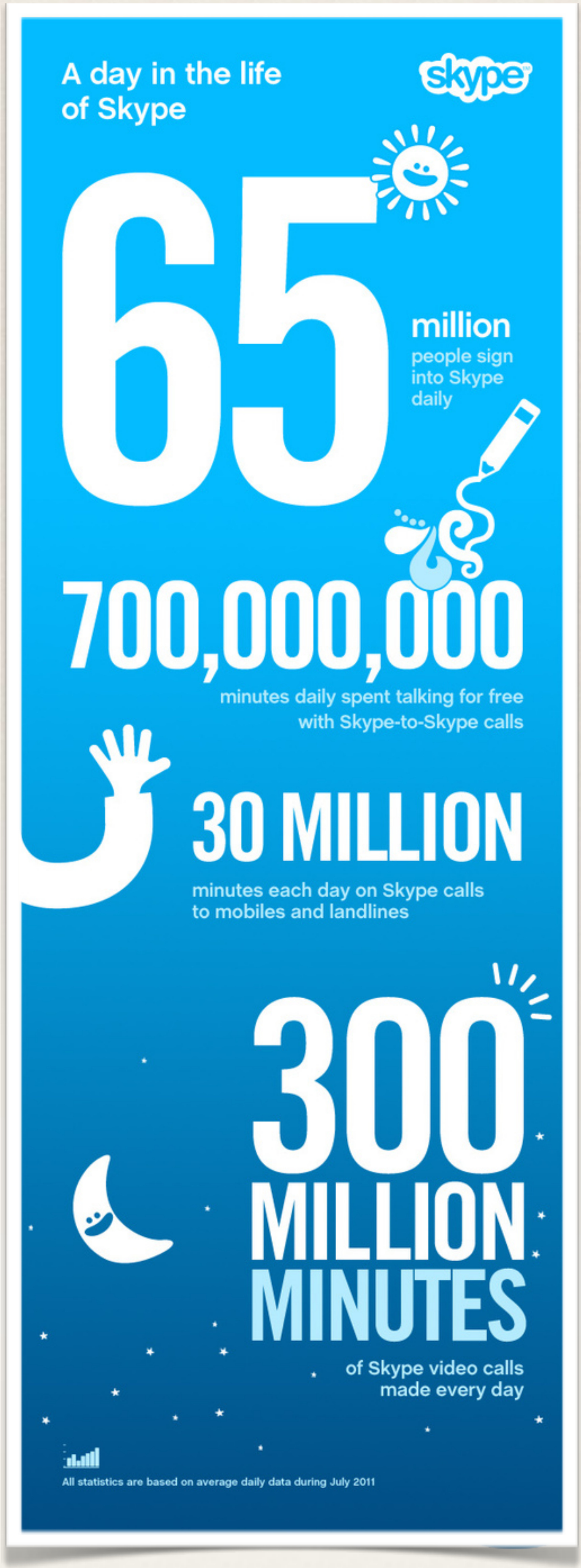
Learning to Read Encrypted VoIP Conversations

Fabian Monrose

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Voice over IP (VoIP)

- Popular replacement for traditional telephony
- Many free, or inexpensive, services available
  - very reliable
  - easy to use



A day in the life of Skype

skype

65 million people sign into Skype daily

700,000,000 minutes daily spent talking for free with Skype-to-Skype calls

30 MILLION minutes each day on Skype calls to mobiles and landlines

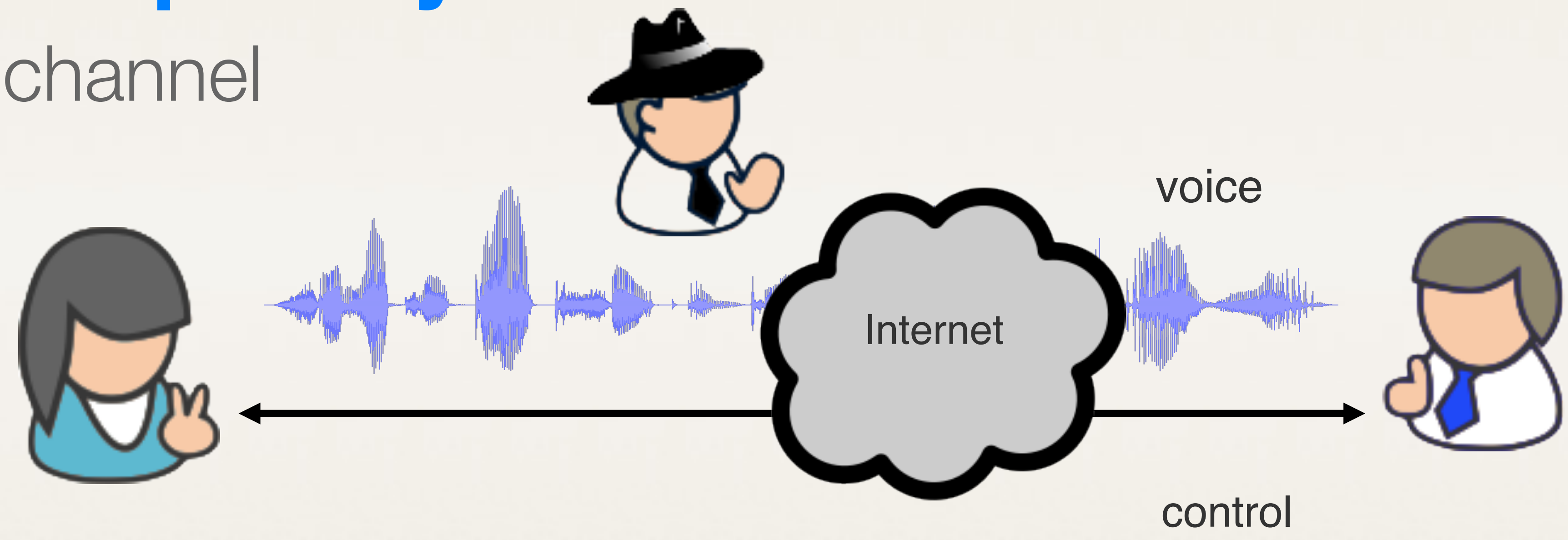
300 MILLION MINUTES of Skype video calls made every day

All statistics are based on average daily data during July 2011

VoIP Security

- Security and privacy implications still not well understood
- Two channels: ***voice*** and ***control***
- Majority of security analyses focus on control channel
  - *e.g., caller id spoofing, registration hijacking, denial of service*

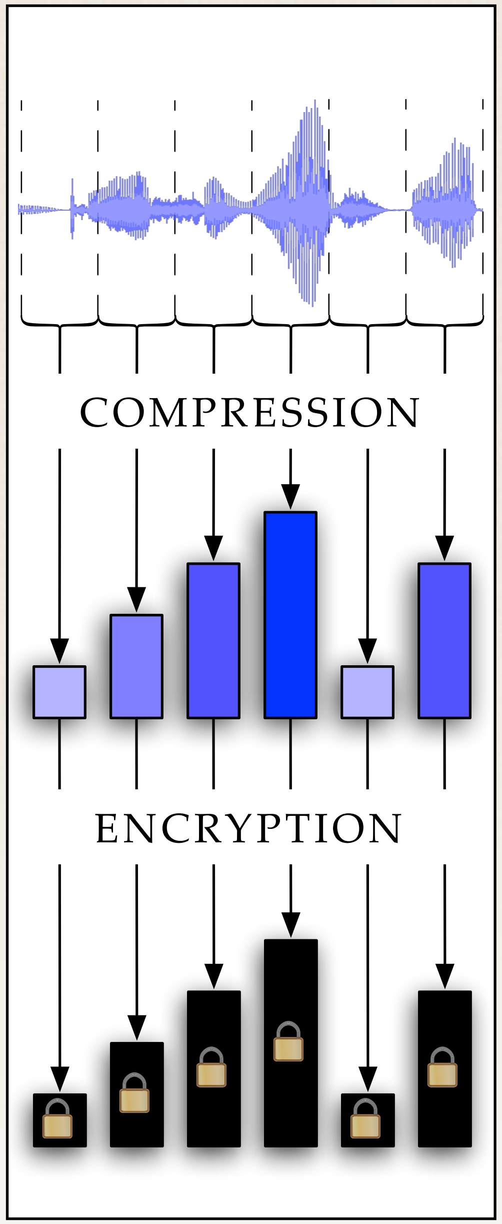
We are interested in the **privacy** of the **voice** channel



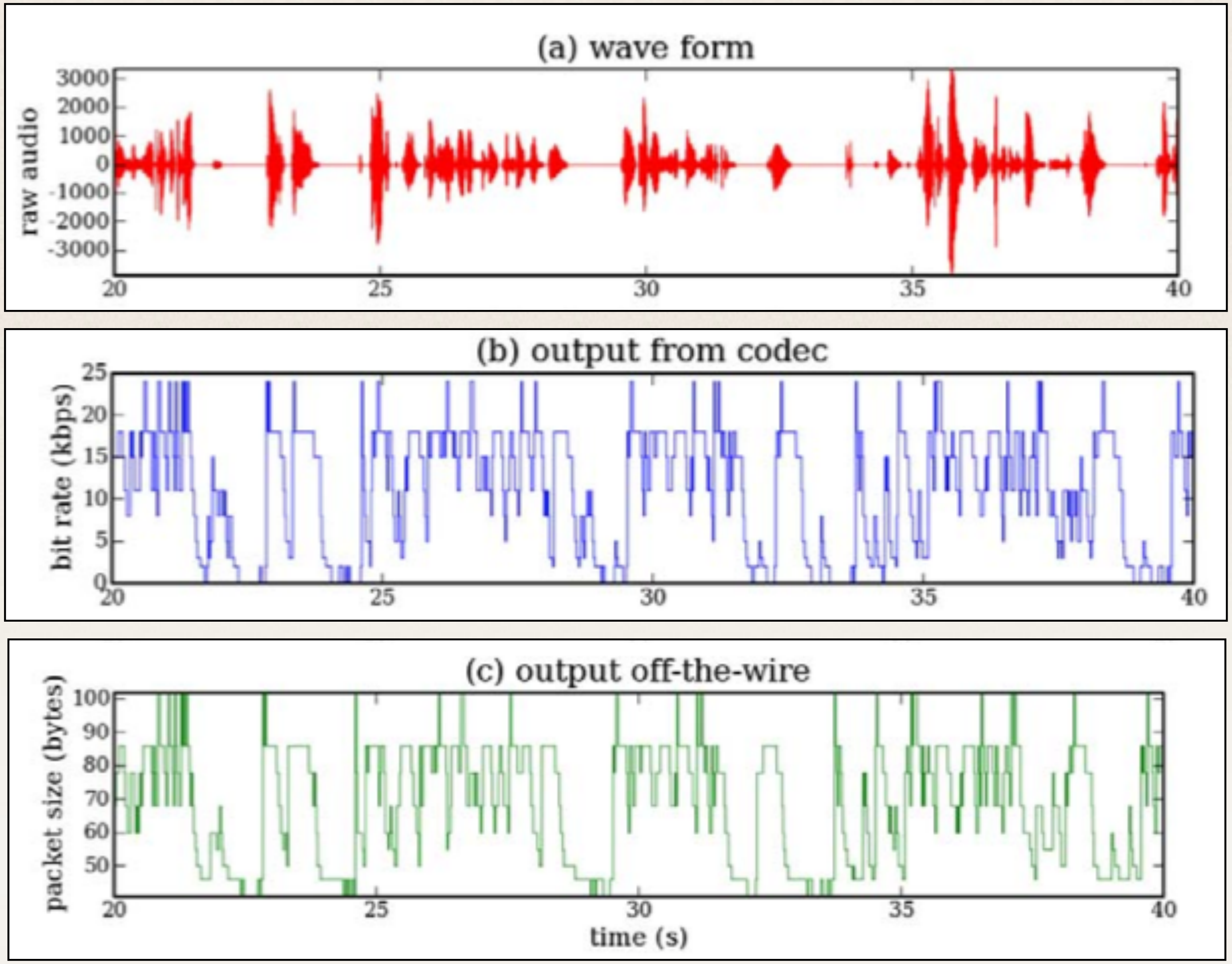
Information leakage

Overlooked interaction of two design decisions:

- **compression**: **variable-bit-rate** (VBR) codecs
  - compress different sounds with varying fidelity
- **encryption**: **length-preserving** stream ciphers



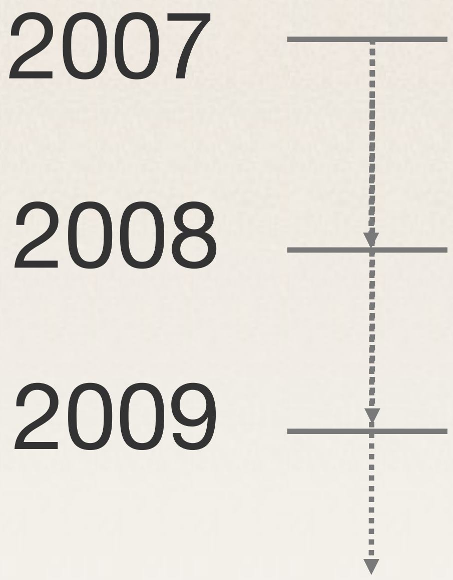
Information leakage



Result: packet sizes reflect properties of the input signal

How bad is this leak?

- Sufficient to determine:



2007

- Wright et al.; Language identification of encrypted VoIP traffic: ***Alejandra y Roberto or Alice and Bob?***, USENIX Security

2008

- Wright et al., *Spot me if you can:* ***Uncovering spoken phrases*** *in encrypted VoIP conversations*, IEEE S&P

2009

- Backes et al.; ***Speaker recognition*** *in encrypted VoIP streams*, ESORICS, 2009.

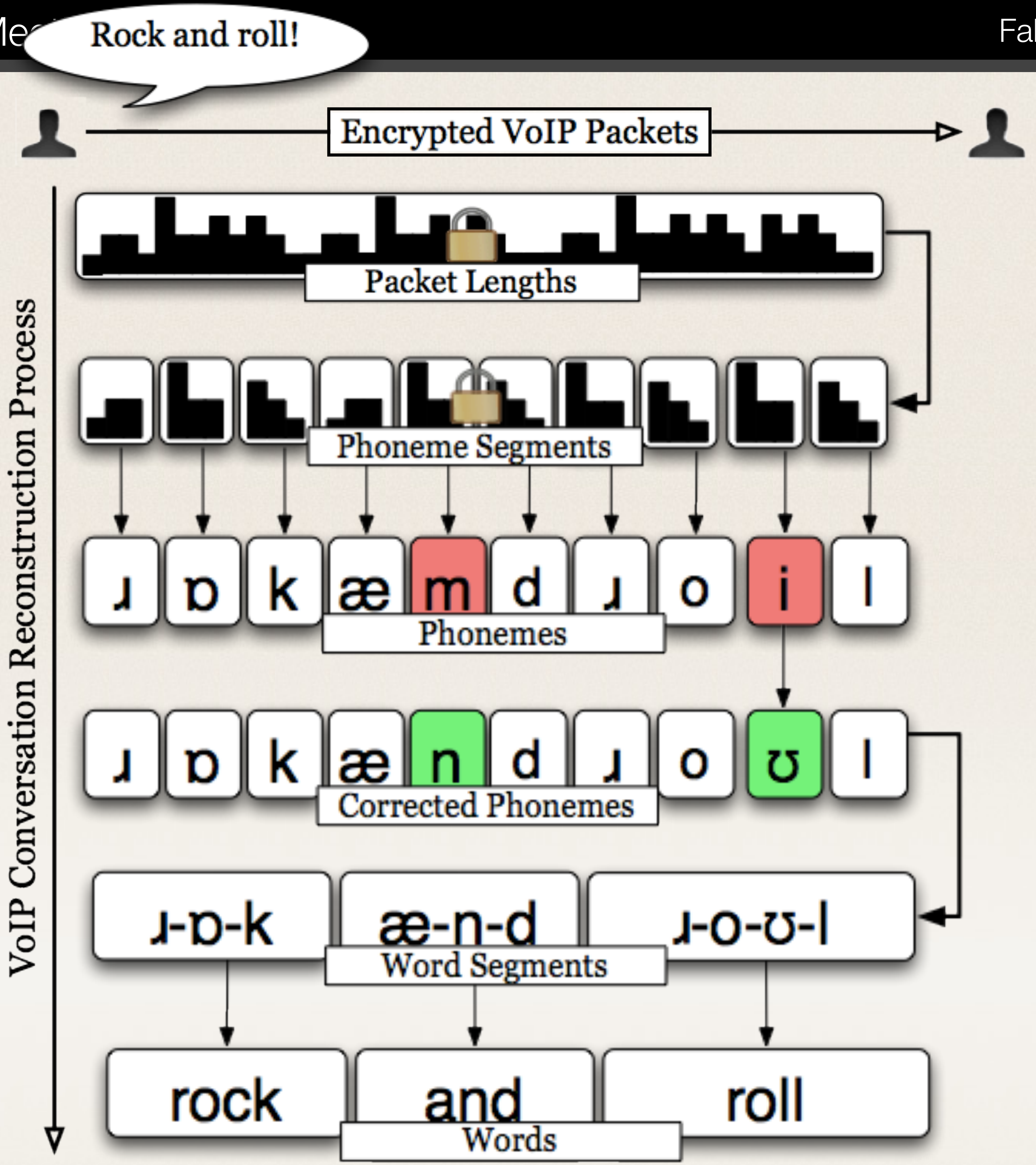
Prior work did not take advantage of language-specific constraints or permitted sequences (i.e., "**phonotactics**")

Mommy Speech Therapy

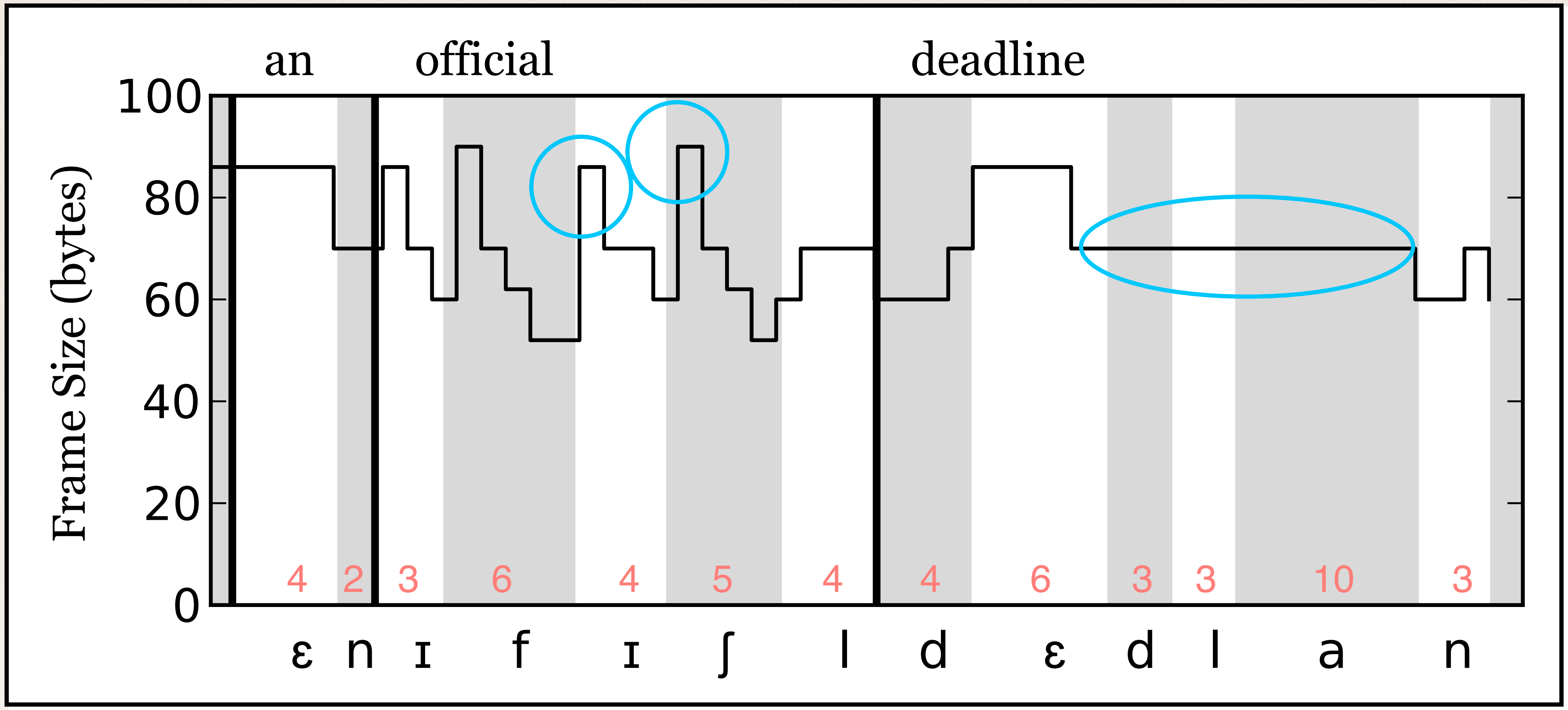
- Infants use perceptual, social, and **linguistic** cues to segment the stream of sounds
- use learned knowledge of **well-formedness**
  - amazingly, infants learn these rudimentary constraints while simultaneously segmenting words
- use familiar words (e.g., their own name, “mama,” etc) to identify new words in a stream

Blanchard et al. *Modeling the contribution of phonotactic cues to the problem of word segmentation*. **Journal of Child Language**, 2010.

Bortfeld et al. ***Mommy and me****: Familiar names help launch babies into speech-stream segmentation*. **Psychological Science**, 2005.


Rock and roll!
Encrypted VoIP Packets
VoIP Conversation Reconstruction Process
Packet Lengths
Phoneme Segments
ɹ ɒ k æ m d ɹ o i l
Phonemes
ɹ ɒ k æ n d ɹ o ʊ l
Corrected Phonemes
ɹ-ɒ-k
æ-n-d
ɹ-o-ʊ-l
Word Segments
rock
and
roll
Words

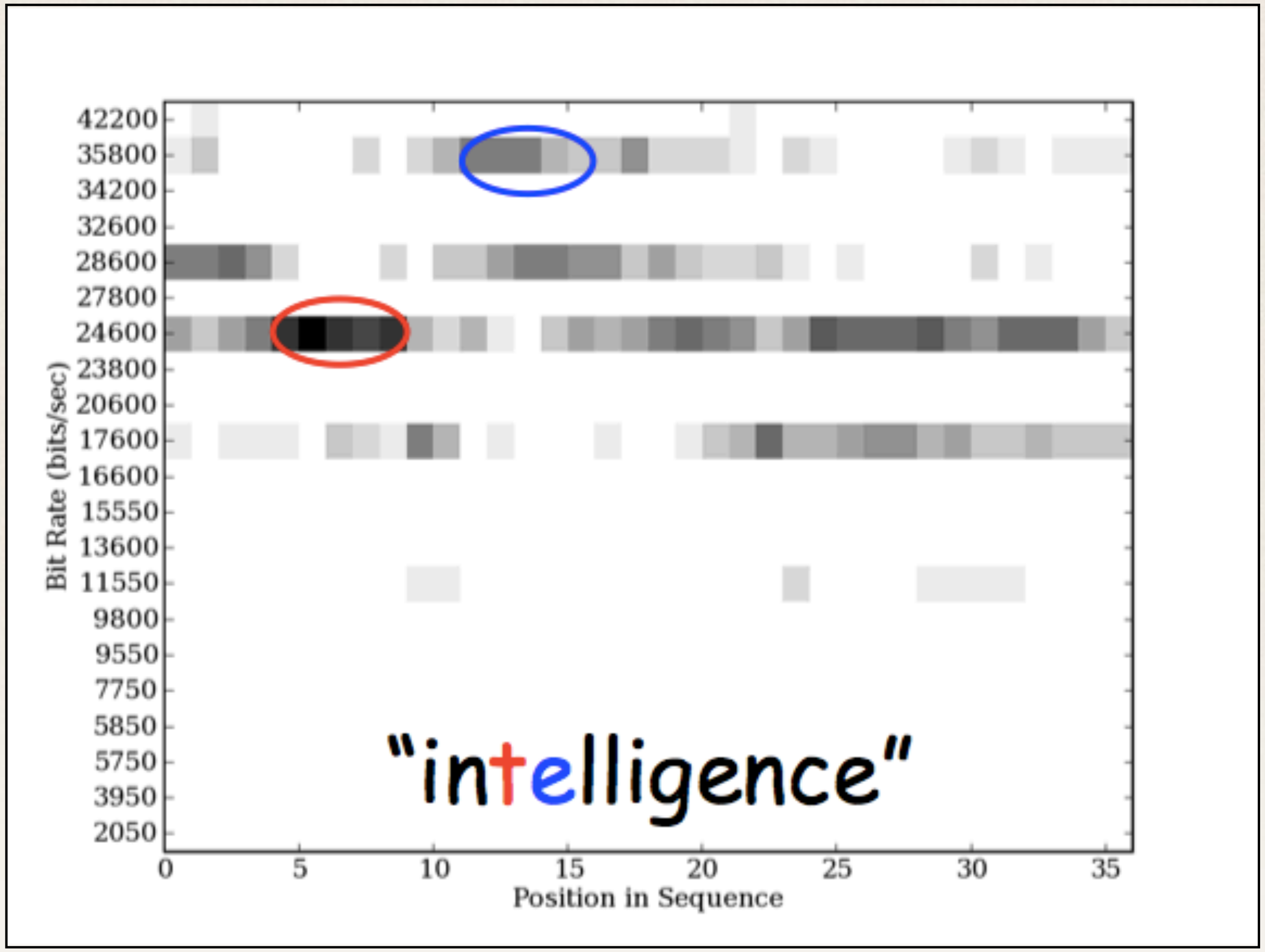
Step 1: phonetic segmentation

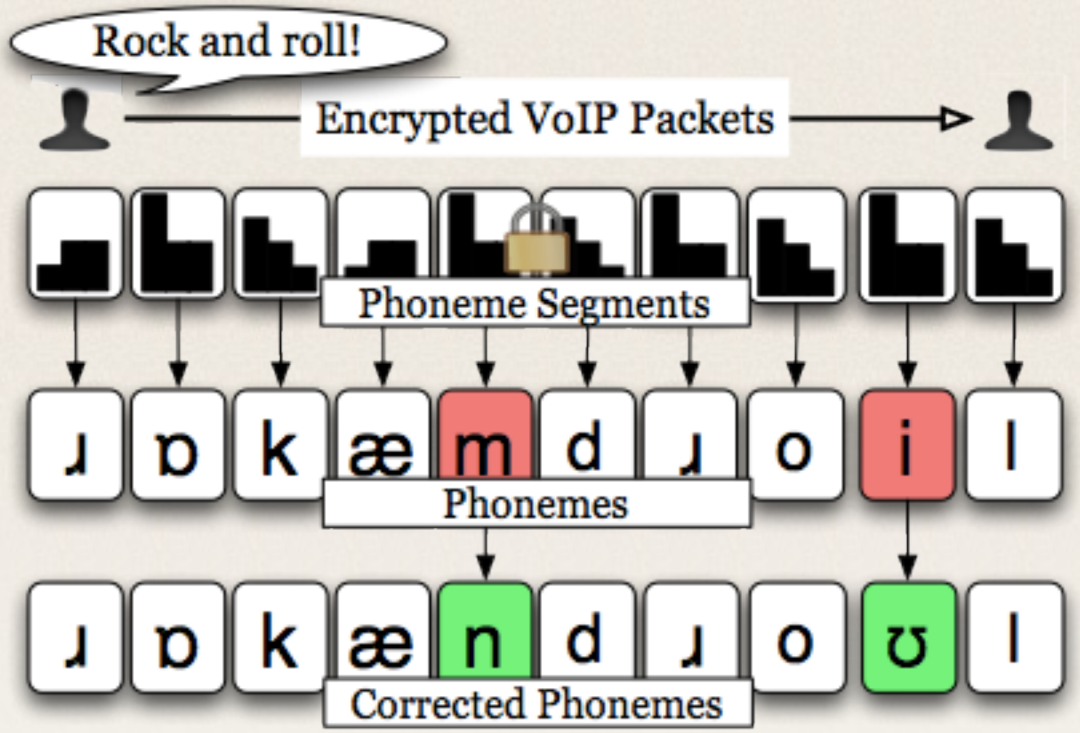


IPA Pronunciation of the phrase “an official deadline”

Observation: frame sizes differ in response to **phoneme transitions**

Step 2: phoneme classification

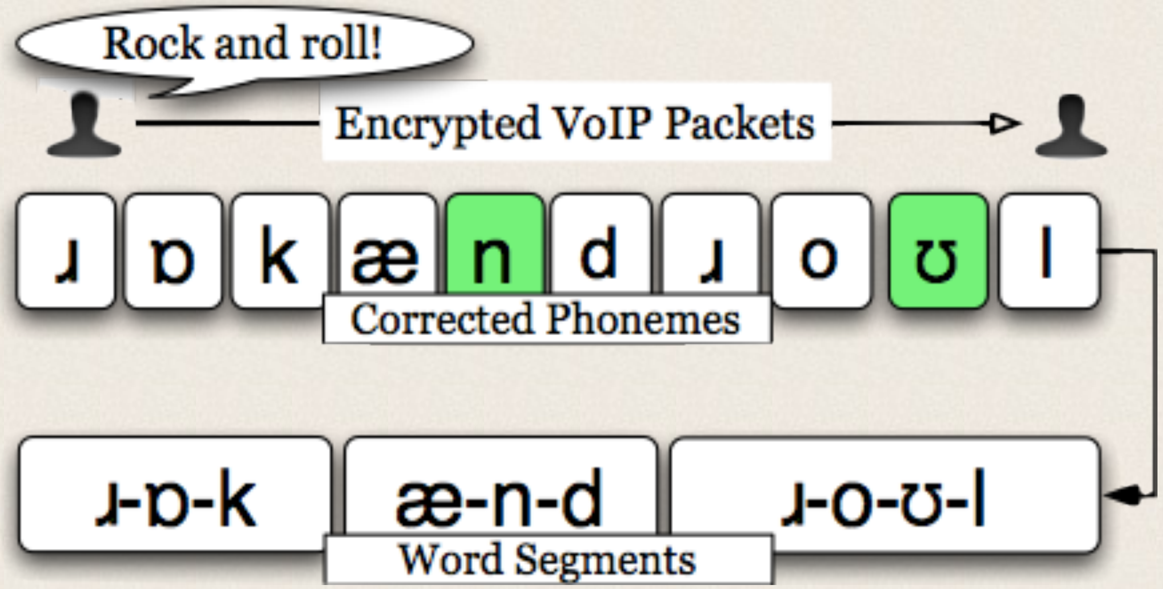




Observation: differing sounds are **encoded** at different bit rates (e.g., **Speex** codec only uses 9 different bit rates in narrow band mode; 21 bit rates in wide-band mode)

Step 3: Word break insertion

Based on language-specific constraints on **phoneme order**



- insert potential word breaks into **impossible** phonetic triplets

✦ [ɪŋw] (‘blessing way’)

- resolve **invalid** word beginning / endings

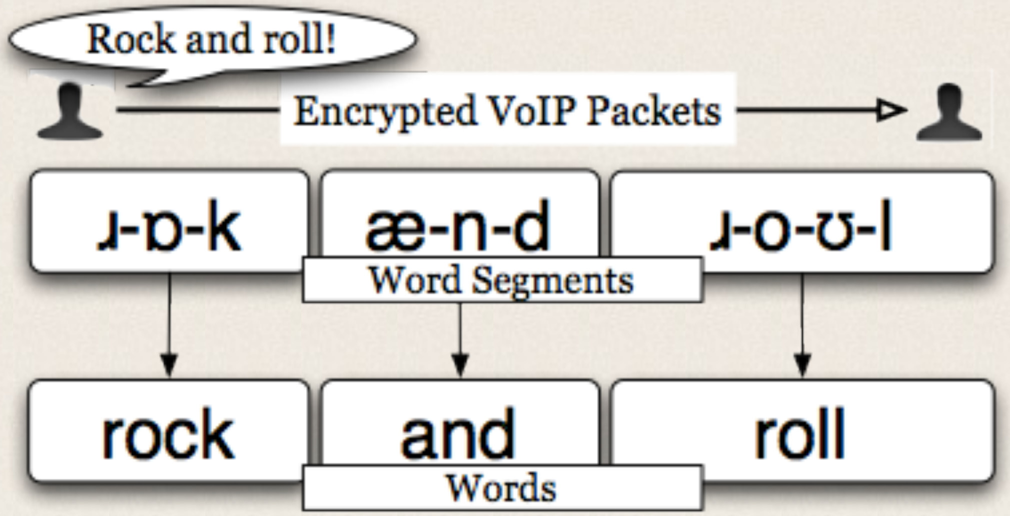
✦ [zdr] (‘eavesdrop’)

- improvement: split resulting segments by **dictionary search**

Harrington et al. Word boundary identification from phoneme sequence constraints in automatic continuous speech recognition. **Computational Linguistics**, **1988**.

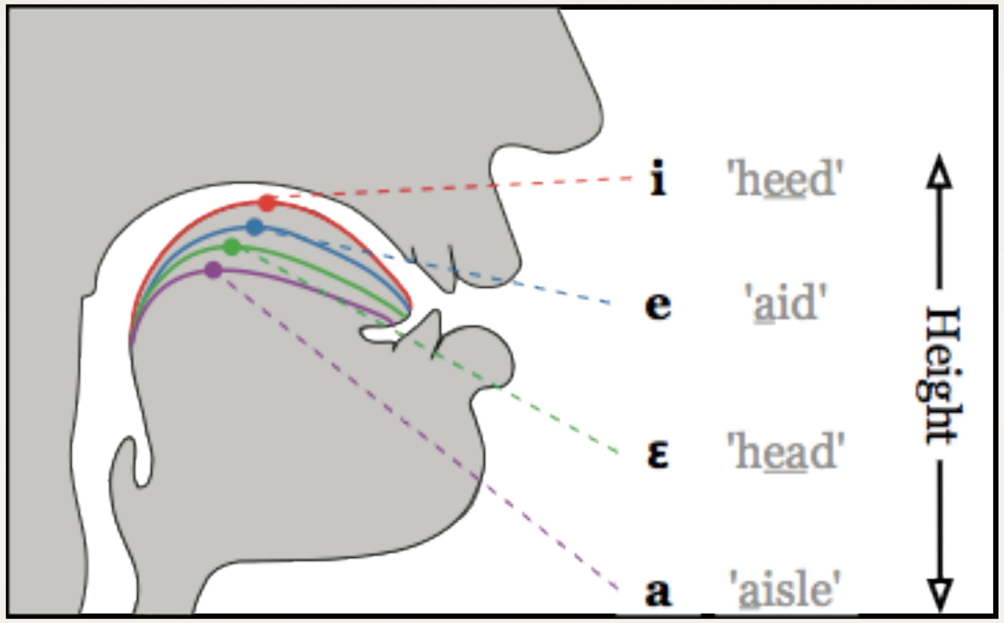
Stage 4: Word Matching

- Find **closest** **pronunciation** using an **edit distance** approach to infer **articulatory** distance between phonemes



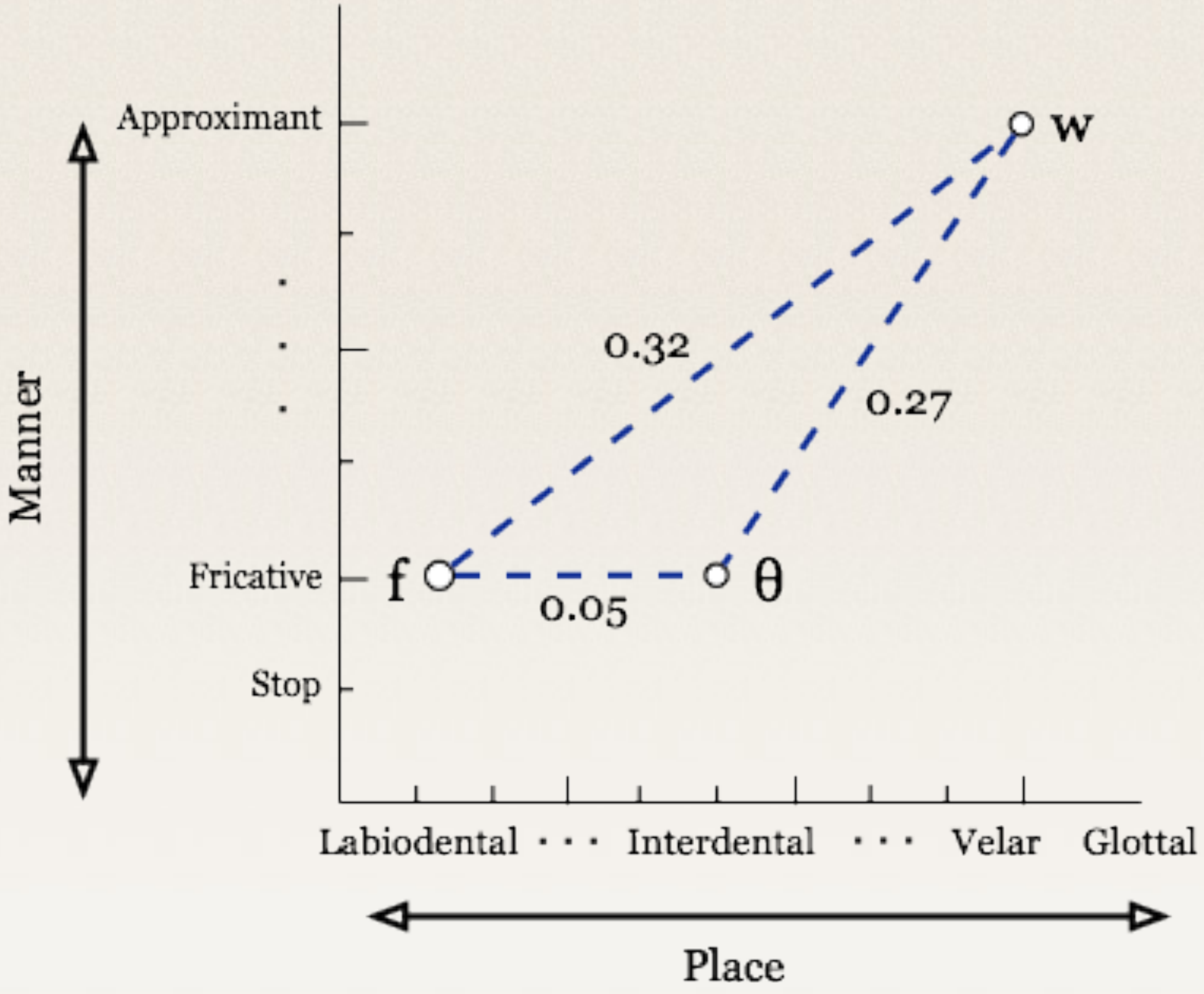
Vowels characterized by tongue position and lip shape (**height, backness, rounding**)

Consonants characterized by restriction of airflow (**place, manner)**



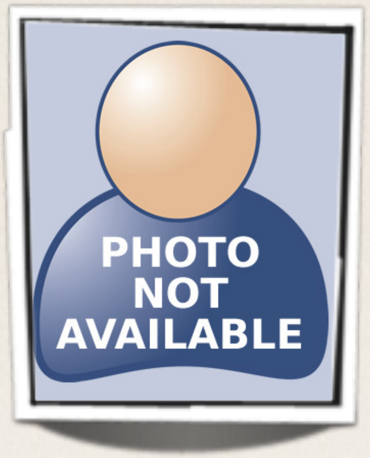
Stage 4: Word Matching

(Or, how we spent the summer of 2011)



Katherine Shaw

Elliott Moreton

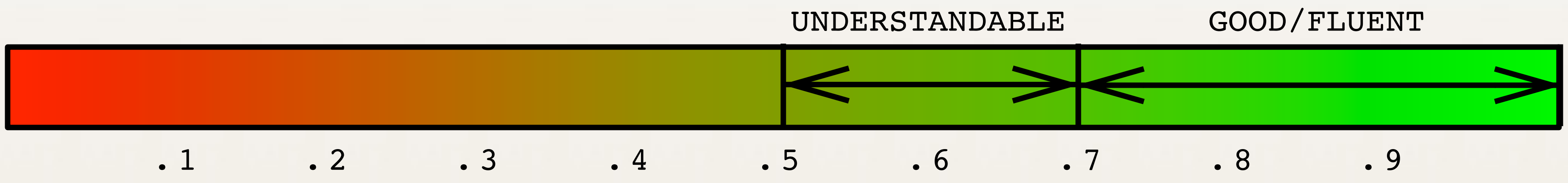


Austin Matthews

Phonetic Edit Distance

Evaluation

- **630** speakers, **8** major dialects of American English
- Score hypotheses using well-studied techniques for modeling the **adequacy** and **fluency** of a translation
- penalizes fragmentation by matching contiguous subsequences (i.e., **fluency**)



METEOR Score Interpretation (Lavie, 2010)

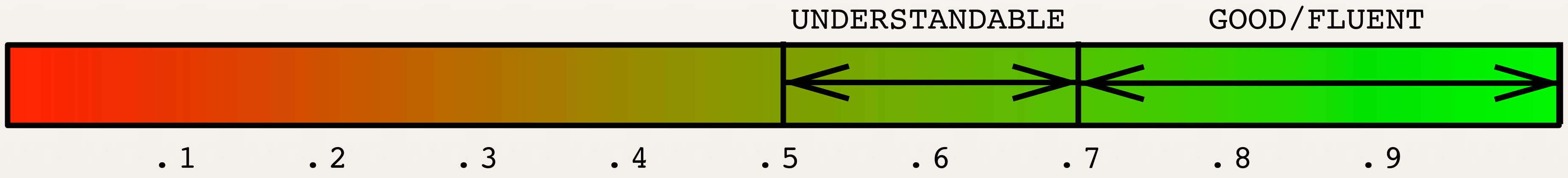
Hypotheses

SA2: "Don't ask me to carry an oily rag like that"	score
Don't asked me to carry an oily rag like that	0.98
Don't ask me to carry an oily rag like dark	0.82
Don't asked me to carry and oily rag like dark	0.80

Context dependent results

Reference **Hypothesis**	score
Change involves the displacement of form. Codes involves the displacement of aim.	0.57
Artificial intelligence is for real. Artificial intelligence is carry all.	0.49
Bitter unreasoning jealousy. Bitter unreasoning dignity.	0.47

Context independent results



METEOR Score Interpretation (Lavie, 2010)

Summary

credit: W. Diffie, S. Landau

- VoIP is here to stay. But, security and privacy issues should not be overlooked
  - quality of reconstructed transcripts better than expected
  - will improve with advancements in computational linguistics
- We need stronger, **interdisciplinary**, partnerships in order to design more secure and efficient solutions

See: A. White, K. Snow, A. Matthews, F. Monrose. Phonotactic Reconstruction of Encrypted VoIP Conversations: hʊkt ɔn fɒnɪks. **IEEE Symposium on Security & Privacy**, 2011.

Ongoing Partnerships

- Closer partnership with Linguistics Department
    - exploring new ways of computing **phonotactic probability** (w/ Elliott Moreton, Katherine Shaw, Jennifer Smith, Andrew White)
    - Linguists are interested in generating and rating new "**blends**"; many applications in Computer Security
- Great learning experience!
    - English is far more complex than I ever imagined
        - e.g., differences in written and spoken form (codas, onsets, nuclei, rhyme, etc.)
- **Strikingly** different lab culture and research meeting practices