# A Novel Approach Using Modulation Features for Multiphone-Based Speech Recognition

**Pascal Clark**, Gregory Sell** and Les Atlas*

---

**Problem Statement**

**Goal:** To extract features for Automatic Speech Recognition (ASR) which are:
1. Low-dimensional (small number of frequency subbands)
2. Low-variance (temporally smooth)

**Our Approach:** Recast the acoustic front-end decomposition as a sum-of-products modulation formula.

\[
x[n] = \sum_{k=0}^{K-1} m_k[n] \cdot c_k[n]
\]

- **Speech:** \( x[n] \)
- **Carrier:** \( c_k[n] \)
- **Modulator:** \( m_k[n] \)

\( a = \) subband number \( \text{and} \) \( n = \) time index

**Key point:** Even within a single subband, the product model is under-determined. (Factoring any number \( c = ab \) has infinite solutions for \( a \) and \( b \).)

**Our contribution:** Two methods of constrained demodulation that estimate bandlimited ASR features \( m_k[n] \) from speech subband signals.

---

**Conventional Demodulation (Hilbert Envelope)**

One row of a spectrogram is the magnitude of an analytic subband signal, or Hilbert envelope. For broad subbands, the Hilbert envelope is band-unlimited and plagued by high-frequency interference. Despite these undesirable traits, the Hilbert envelope still underpins most speech recognition features, such as commonly-used MFCC representations.

---

**New Method 1 – Convex Demodulation**

**Constrained Demodulation is a Starting Point for Speech Recognition Features**

**New Method 2 – Coherent Demodulation**

**New Method 2 – Coherent Demodulation**

**Results and Conclusion**

- **Time-frequency ASR features are related to an under-determined product signal model.**
- **We offer two bandwidth-constrained demodulation methods: convex and coherent, as alternatives to the conventional Hilbert envelope.**
- **Results:**
  - Improved performance in individual multiphone classification (Convex vs. Hilbert, two panels to the left)
  - Slight improvement in word-error rate using Convex or Coherent: ~0.1% relative to Hilbert on 400+ hours of Broadcast News corpus.

We acknowledge Geoffrey Zweig, Patrick Nguyen, and Damianos Karakos for their expertise and contributions.