# Experiment 6

# Robert V Shannon

Research Professor, USC Otolaryngology
Research Professor, USC Biomedical Engineering
Adjunct Professor, USC Neuroscience

## Research Topics

1. Cochlear Implants
2. Auditroy Brainstem Implants
3. Auditory Midbrain Implants
4. Psychophysics of normal and impaired hearing
5. Speech perception with degraded sensory information

## Research Overview

I am interested in how auditory information is coded in the nervous system. ...

https://science.sciencemag.org/content/270/5234/303



REPORTS

## Speech Recognition with Primarily Temporal Cues

Robert V. Shannon[1], Fan-Gang Zeng, Vivek Kamath, John Wygonski, Michael Ekelid
+ See all authors and affiliations

*Science* 13 Oct 1995:
Vol. 270, Issue 5234, pp. 303-304
DOI: 10.1126/science.270.5234.303



### Abstract

Nearly perfect speech recognition was observed under conditions of greatly reduced spectral information. Temporal envelopes of speech were extracted from broad frequency bands and were used to modulate noises of the same bandwidths. This manipulation preserved temporal envelope cues in each band but restricted the listener to severely degraded information on the distribution of spectral energy. The identification of consonants, vowels, and words in simple sentences improved markedly as the number of bands increased; high speech recognition performance was obtained with only three bands of modulated noise. Thus, the presentation of a dynamic temporal pattern in only a few broad spectral regions is sufficient for the recognition of speech.

# Abstract


- Nearly perfect speech recognition was observed under conditions of greatly reduced spectral information.
- Temporal envelopes of speech were extracted from broad frequency bands and were used to modulate noises of the same bandwidths.
- This manipulation preserved temporal envelope cues in each band but restricted the listener to severely degraded information on the distribution of spectral energy.
- The identification of consonants, vowels, and words in simple sentences improved markedly as the number of bands increased; high speech recognition performance was obtained with only three bands of modulated noise.
- Thus, the presentation of a dynamic temporal pattern in only a few broad spectral regions is sufficient for the recognition of speech.

# Steps

- The acoustic signal was divided into several frequency bands and the amplitude envelope was extracted from each band by half-wave rectification and low-pass filtering.
- Low-pass filters with cutoff frequencies of 16, 50, 160, and 500 Hz were used for envelooe extraction to evaluate the effect of reducing the bandwidth of temporal envelope information.
- The envelope signal was used to modulate white noise, which was then spectrally limited by the same bandpass filter used for the original analysis band.

# Steps (Continued)

- All bands were then summed and presented to the listeners through headphones.
- One, two, three, or four band processors were used, each with envelope information low-pass-filtered at 16, 50, 160, or 500 Hz, for a total of 16 conditions.
- The combined signal was low-pass-filtered at 4 kHz, amplified and presented to the listener through headphones.
- The listeners were instructed to identify the presented stimulus by selecting it from the complete set of 16 consonants or 8 vowels. Sentences were presented once and the listeners were instructed to repeat as many words as they could.

# Steps (continued)

- The audio signal was digitized at a 10-kHz sampling rate. The signal was then split into frequency bands. Adjacent filters overlapped at the point at which the output from each filter was 15 dB down from the level in the pass-band.
- Filter cutoff frequencies were 1500 Hz for the two-band processor, 800 and 1500 Hz for the three-band processor, and 800, 1500, and 2500 Hz for the four-band processor