Does the Auditory System Efficiently Code All Languages Or Just American English?

PS-181

Christian Stilp and Lily Assgari
Department of Psychological and Brain Sciences, University of Louisville

INTRODUCTION

Lewicki (2002) used Independent Component Analysis (ICA) to examine statistical properties of human speech. Statistically optimal filters for encoding speech were well-aligned with frequency tuning in the mammalian auditory nerve (comparing measures of Q_{10}), leading Lewicki to suggest speech makes efficient use of coding properties of the auditory system. However, these analyses only examined American English, which is neither normative nor representative of the world's languages. Here, ICA was used to compare optimal encoding of speech from 14 different languages found across the world with physiological response properties.

METHODS

Stimuli

Recordings of 14 languages (Dutch, Flemish, Greek, Javanese, Ju|'hoan, Norwegian, Swedish, Tagalog, Tahitian, Urhobo, Vietnamese, Wari', Xhosa, Yeyi) were collected, mostly from the UCLA Phonetics Lab Archive (http://archive.phonetics.ucla.edu/). All recordings were at least one minute long and contained clear speech tokens from a native speaker without any background noise. Recordings were high-pass filtered at 125 Hz and divided into 8-ms samples (after Lewicki, 2002).

ICA

In ICA, the observed data **x** are assumed to be the result of linear combinations of **s**:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \qquad [1]$$

where **A** is a mixing matrix whose columns constitute basis functions, and **s** is a source vector with components s_i that are statistically independent from each other. **A** and **s** are unknown, so ICA estimates them as follows:

$$\mathbf{y} = \mathbf{W}\mathbf{x} \qquad [2]$$

W is an unmixing matrix of the same dimensionality as **A** ($\mathbf{W} = \mathbf{A}^{-1}$), making the output **y** the recovered source vector which approximates **s** up to scaling and permutation. The rows of **W** are statistically optimal filters for recovering source signals **s** from the observed mixtures **x**.

Maximum likelihood ICA was used (Pearlmutter & Parra, 1996) with the natural gradient extension to facilitate convergence. **W** was iteratively updated by stochastic gradient descent:

$$\Delta\mathbf{W} = [\mathbf{I} - \mathrm{sign}(\mathbf{y})\mathbf{y}^{T}]\mathbf{W} \qquad [3]$$

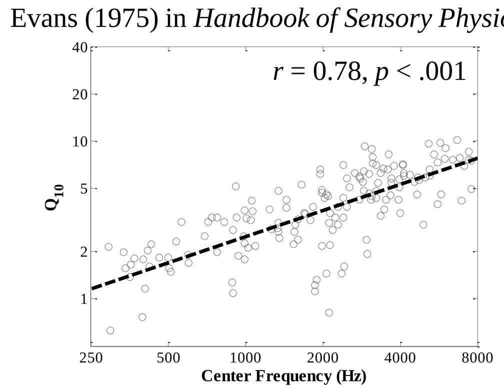
where **I** is the identity function and sign(·) is the sign function. **W** is initialized to the identity matrix, and Δ**W** is the change in the unmixing matrix that is added to **W** at each iteration. ICA was conducted for 20,000 iterations, with a different batch of 500 samples randomly selected for analysis at each iteration.

Regression Analysis

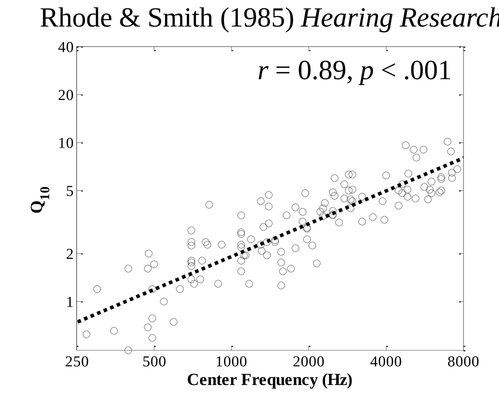
When ICA is complete, each row in **W** is a statistically optimal filter for encoding input stimuli. Sharpness of each filter (Q_{10}) was calculated when possible (when filter response decreased by 10 dB above and below the center frequency). Linear regressions were calculated for Q_{10} as a function of center frequency on a log-log scale, following Lewicki (2002).

The center frequency (up to 8 kHz) and sharpness (Q_{10}) of auditory nerve fibers in cats show highly linear relationships. The two examples used by Lewicki (2002) are shown below, with linear regression fits superimposed. Each circle represents one tuning curve.

Evans (1975) in *Handbook of Sensory Physiology*



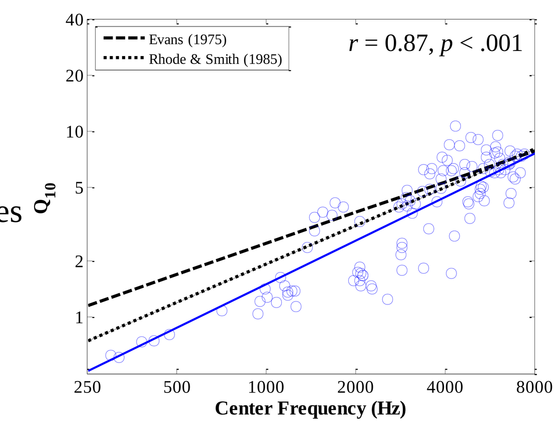
Rhode & Smith (1985) *Hearing Research*



ICA produces statistically optimal filters for encoding a set of sounds. Lewicki (2002) reported that optimal filters for encoding American English were a good match for these physiological measures. How does this relationship hold for other languages?

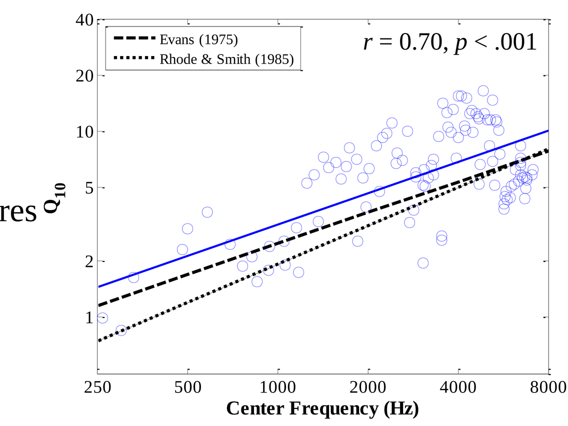
Dutch
Regions: Netherlands, North Belgium (see Flemish), Netherlands Antilles, Aruba, Suriname
Family: West Germanic

Excellent fit to physiological measures



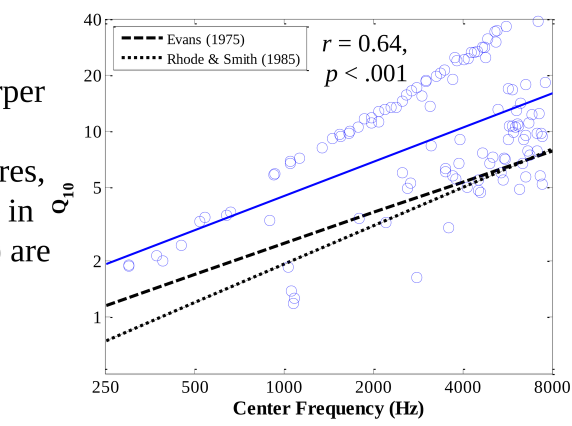
Flemish
Regions: North Belgium (Dutch dialect)
Family: West Germanic

Excellent fit to physiological measures



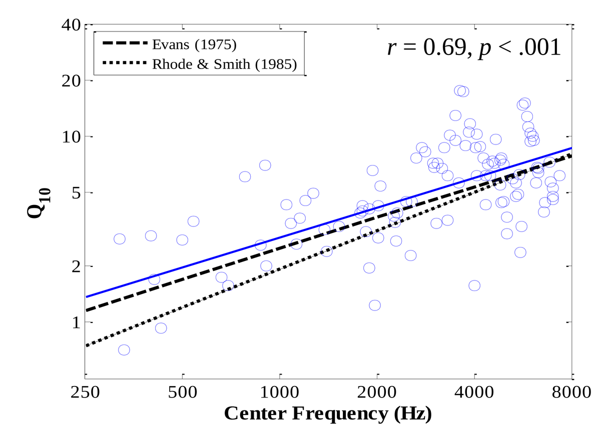
Greek
Regions: Greece, regions all over the world
Family: Greek/Hellenic

Many filters are sharper (higher Q_{10}) than physiological measures, but slopes (increase in sharpness across CF) are comparable



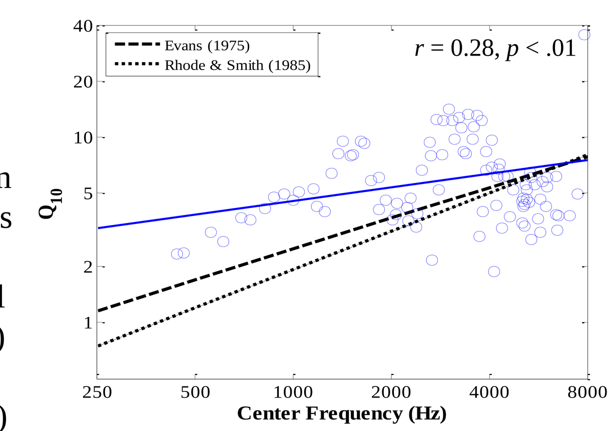
Javanese
Regions: Indonesia, communities in Malaysia, Suriname, New Caledonia, Netherlands
Family: Western Malayo-Polynesian brand of the Austronesian languages

Excellent fit to physiological measures



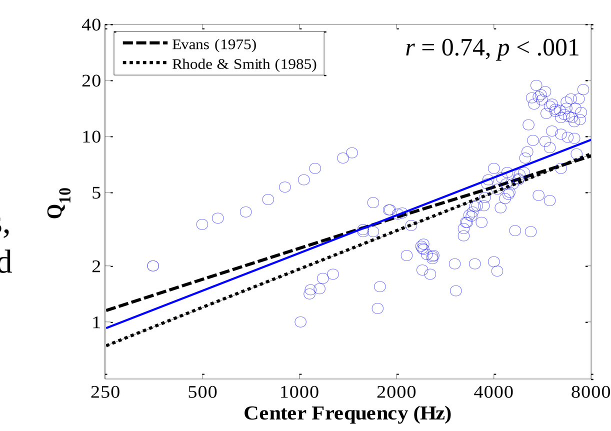
Ju'|hoan
Regions: Botswana, Namibia
Family: Khoisan Language, !Kung Family

Shallow slope suggests broader frequency resolution and finer temporal resolution in filters, suitable for highly transient clicks in Ju'|hoan. Shallow slopes were also reported for transient environmental sounds (Lewicki, 2002) and stop consonants (Stilp & Lewicki, 2014)



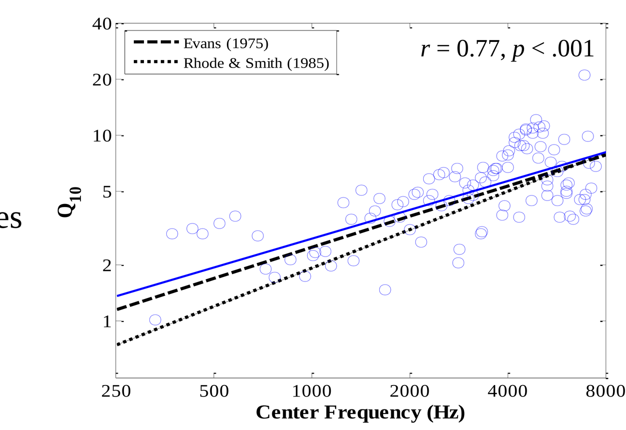
Norwegian
Regions: Norway
Family: North Germanic

A subset of low-frequency filters are sharper than physiological measures, but otherwise very good fit



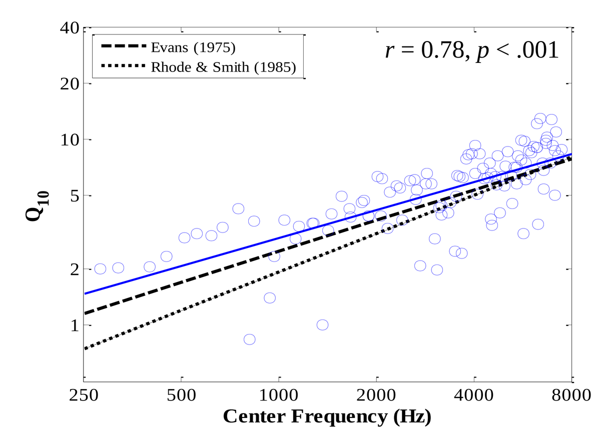
Swedish
Regions: Sweden, parts of Finland
Family: North Germanic

Excellent fit to physiological measures



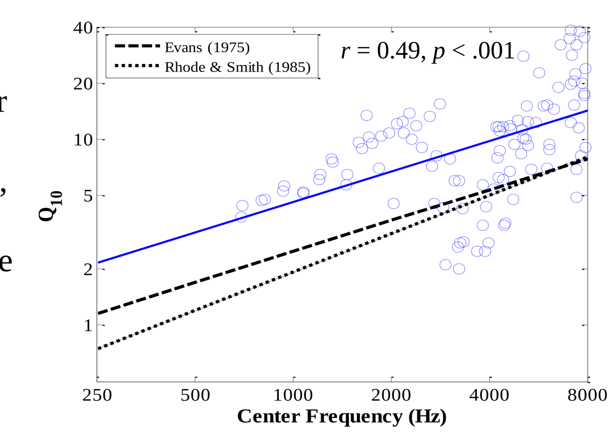
Tagalog
Regions: Republic of Philippines
Family: Central Philippine group of the Philippine subgroup of the Western-Malayo-Polynesian branch of the Malayo-Polynesia subfamily of the Austronesian language family

Excellent fit to physiological measures



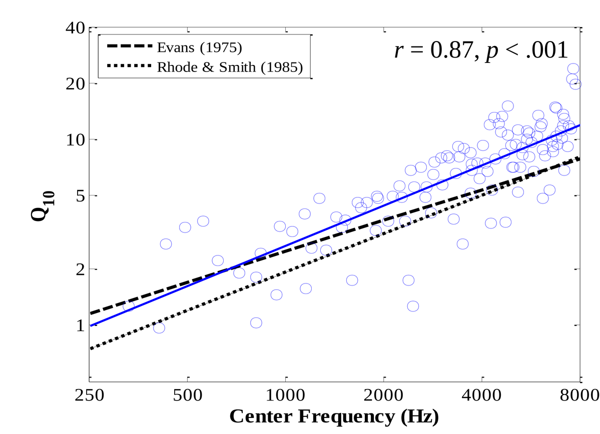
Tahitian
Regions: Polynesian Triangle, Tahiti
Family: Polynesian Languages, Austronesian

Many filters are sharper (higher Q_{10}) than physiological measures, but slopes (increase in sharpness across CF) are comparable



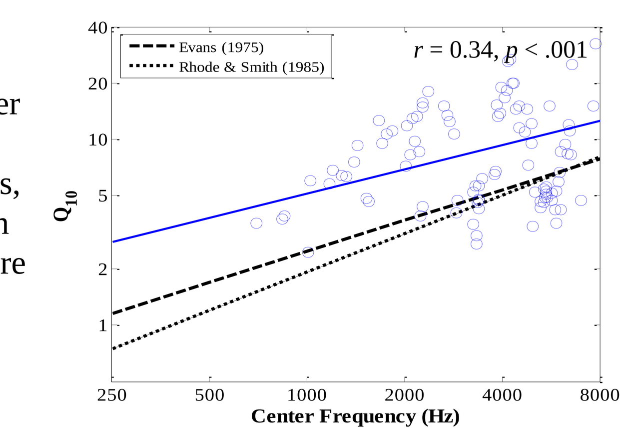
Urhobo
Regions: Nigeria
Family: Niger-Congo

Excellent fit to physiological measures



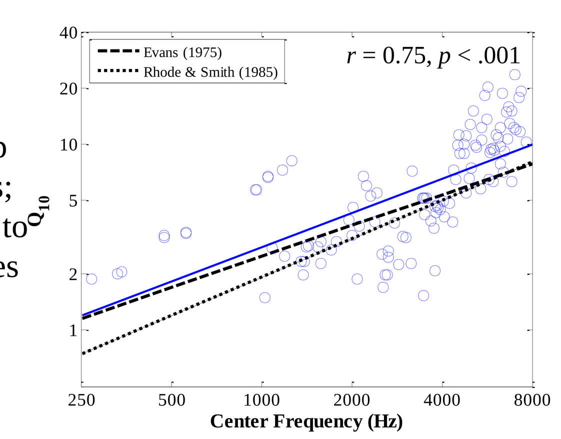
Vietnamese
Regions: Vietnam, Parts of Kampuchea (Cambodia), Thailand, Laos and oversea communities
Family: Muong-Vietnamese subgroup of the Mon-Khmer subfamily of the Austro-Asiatic family

Many filters are sharper (higher Q_{10}) than physiological measures, but slopes (increase in sharpness across CF) are comparable



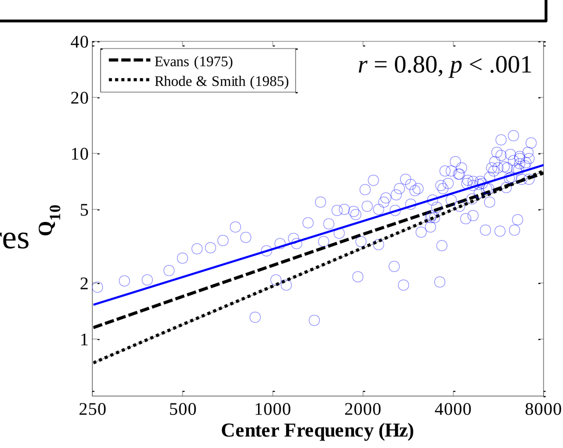
Wari
Regions: Brazil
Family: Chupacura, Madeira

High density of sharp high-frequency filters; otherwise excellent fit to physiological measures



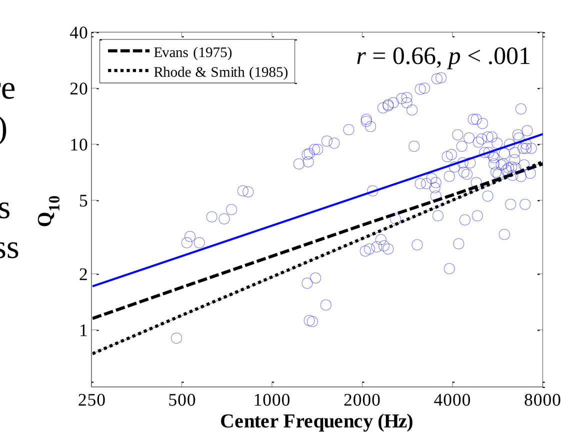
Xhosa
Regions: South West Cape Province and Transkei in the Republic of South Africa
Family: Nguni group of the Bantu sub branch of the Benue-Congo brand of the Niger-Congo subfamily of the Niger-Khordofanian family

Excellent fit to physiological measures



Yeyi
Regions: Northwest Botswana, Namibia, East Caprivi, Ngamilan
Family: Bantu

A subset of filters are sharper (higher Q_{10}) than physiological measures, but slopes (increase in sharpness across CF) are comparable



DISCUSSION

Results extend work by Lewicki (2002), as filters that optimally encode speech sounds in a wide variety of languages (not just American English) generally align with tuning properties in the mammalian auditory nerve. These matches are particularly strong for other Germanic languages (Dutch, Flemish, Norwegian, Swedish). Further research is needed to understand why some languages required sharper filters than those observed physiologically (Greek, Tahitian, Vietnamese), but regression slopes were still comparable. Results support the efficient coding hypothesis (Barlow, 1961), as the auditory system has evolved to optimally encode a wide range of speech sounds across languages.