Languages Across the World are Efficiently Coded by the Auditory System

2pSC21

UNIVERSITY OF LOUISVILLE

Christian Stilp and Ashley 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, 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 revealed optimal encoding of speech from languages found across the world; these were then compared to physiological response properties from the mammalian auditory system.

METHODS

Stimuli

Recordings of 15 languages were collected, mostly from Global Recordings Network (http://globalrecordings.net/). All recordings were roughly 10 minutes long (Tahitian was 7 min.) and contained clear speech tokens without any background noise. Recordings came from multiple talkers whenever possible. Recordings were high-pass filtered at 125 Hz and divided into 8-ms samples (after Lewicki, 2002).

ICA

In ICA, the observed data $\mathbf{x}$ are assumed to be the result of linear combinations of $\mathbf{s}$:

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

where $\mathbf{A}$ is a mixing matrix and $\mathbf{s}$ is a source vector with statistically independent components $\mathbf{s}_i$. $\mathbf{A}$ and $\mathbf{s}$ are unknown, so ICA estimates them as follows:

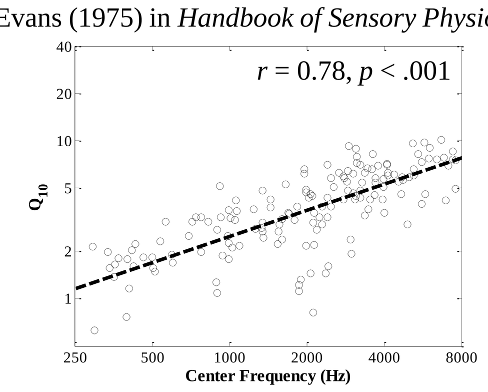
$$\mathbf{y} = \mathbf{W}\mathbf{x} \qquad [2]$$

$\mathbf{W}$ is an unmixing matrix of the same dimensionality as $\mathbf{A}$ ($\mathbf{W} = \mathbf{A}^{-1}$). The rows of $\mathbf{W}$ are statistically optimal filters for recovering source signals $\mathbf{s}$ from the observed mixtures $\mathbf{x}$. Maximum likelihood ICA was used with the natural gradient extension to facilitate convergence. For each language, ICA was conducted for 20,000 iterations, with a different batch of 500 samples randomly selected for analysis at each iteration. For more details, see Stilp and Lewicki (2014 *POMA*).

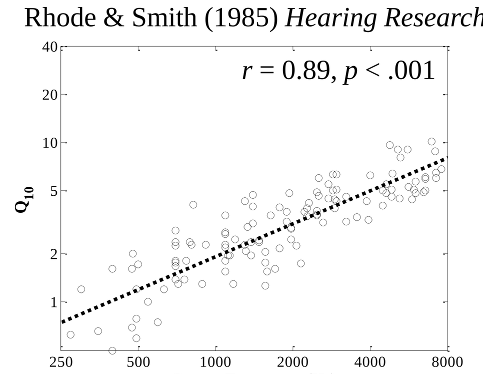
Regression Analysis

The center frequency (up to 8 kHz) and sharpness (Q_{10}; center frequency / bandwidth -10 dB from peak) 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*



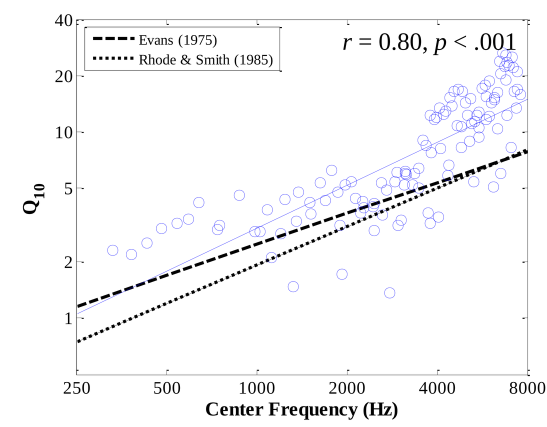
Rhode & Smith (1985) *Hearing Research*



These measures are an excellent fit to statistically optimal filters for encoding American English, but do they fit other languages as well? To answer this question, ICA was conducted on each language. The sharpness of each filter (row in $\mathbf{W}$) was calculated using Q_{10}. Linear regressions were calculated for Q_{10} as a function of center frequency on a log-log scale.

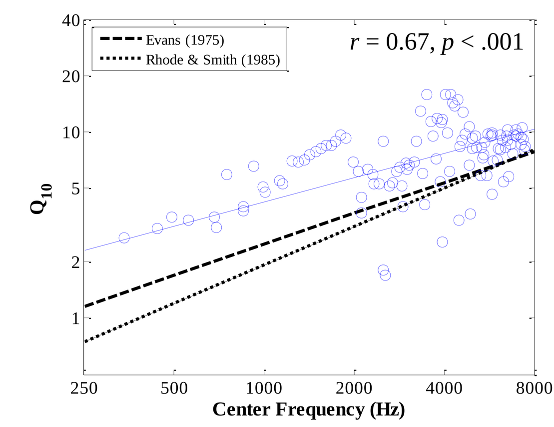
Dutch (1 talker)
Regions: Netherlands, North Belgium (see Flemish), Netherlands Antilles, Aruba, Suriname
Family: West Germanic

Excellent fit to physiological measures



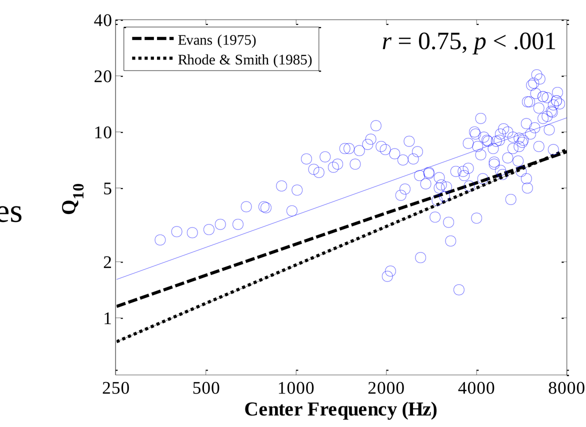
Flemish (6 talkers)
Regions: North Belgium (Dutch dialect)
Family: West Germanic

Filters slightly sharper than physiological measures; otherwise excellent fit



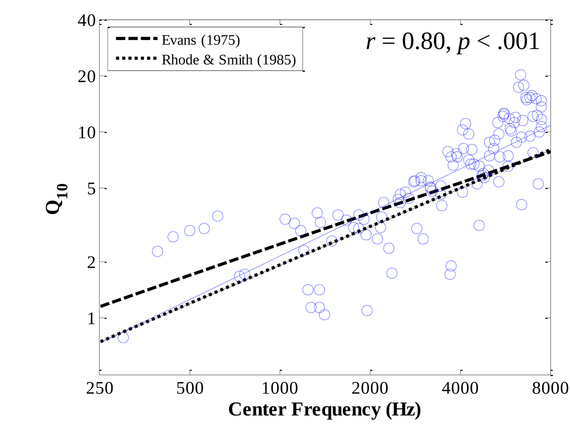
Greek (2 talkers)
Regions: Greece, regions all over the world
Family: Greek/Hellenic

Excellent fit to physiological measures



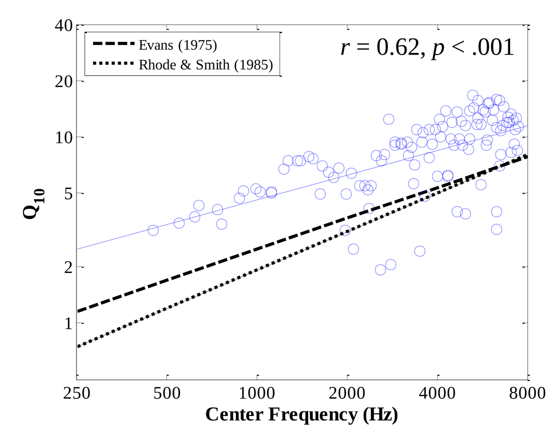
Javanese (1 talker)
Regions: Indonesia, communities in Malaysia, Suriname, New Caledonia, Netherlands
Family: Western Malayo-Polynesian brand of the Austronesian languages

Excellent fit to physiological measures



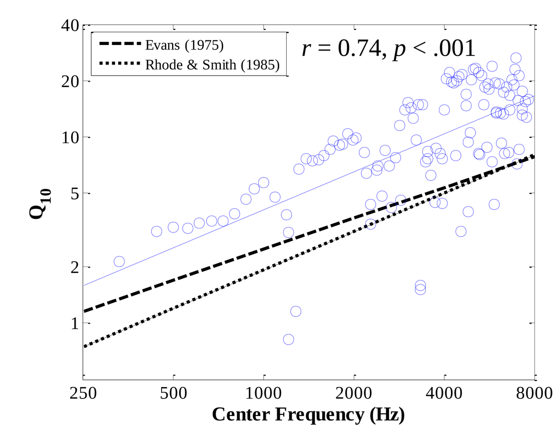
Ju'|hoan (3 talkers)
Regions: Botswana, Namibia
Family: Khoisan Language, !Kung Family

Filters slightly sharper than physiological measures; otherwise excellent fit



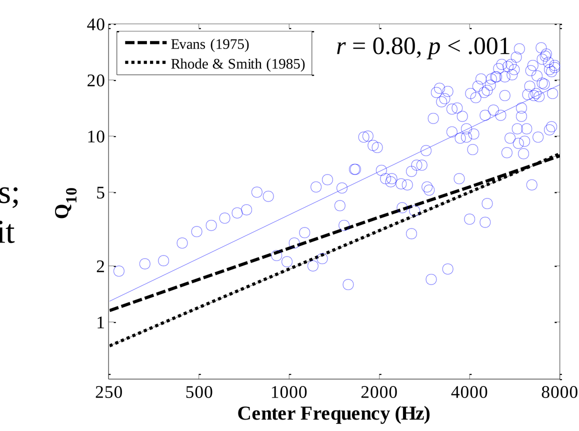
Mandarin Chinese (87 talkers)
Regions: Primarily north of the Yangtze River in China; Taiwan
Family: Sinitic branch of Sino-Tibetan

Filters slightly sharper than physiological measures; otherwise very good fit



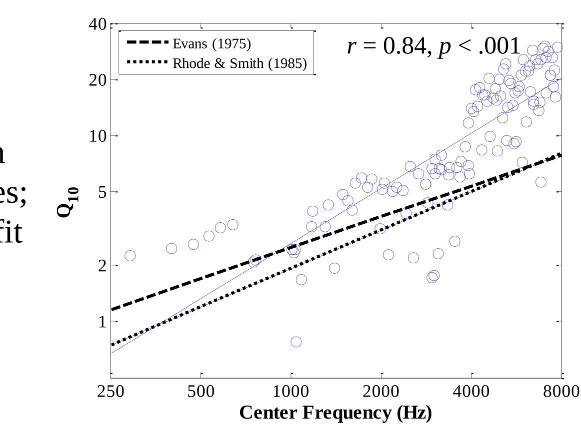
Norwegian (5 talkers)
Regions: Norway
Family: North Germanic

Slope is steeper than physiological measures; otherwise very good fit



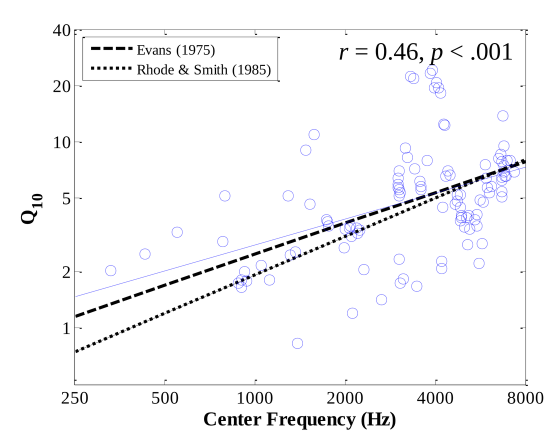
Swedish (1 talker)
Regions: Sweden, parts of Finland
Family: North Germanic

Slope is steeper than physiological measures; otherwise very good fit



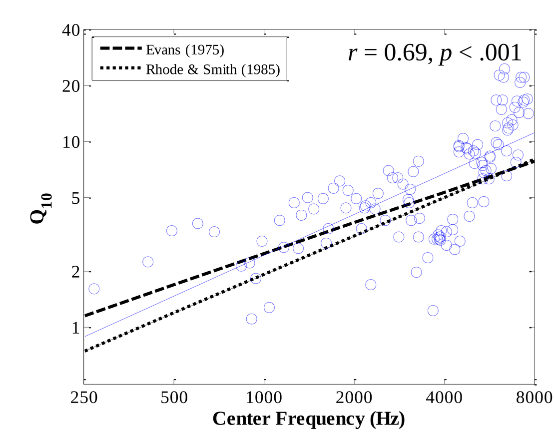
Tagalog (1 talker)
Regions: Republic of Philippines
Family: Central Philippine group of the Philippine subgroup of the Western-Malayo-Polynesian branch of the Malayo-Polynesia subfamily of the Austronesian language family

Filters are highly variable for this talker but regression function is still an excellent fit to physiological measures



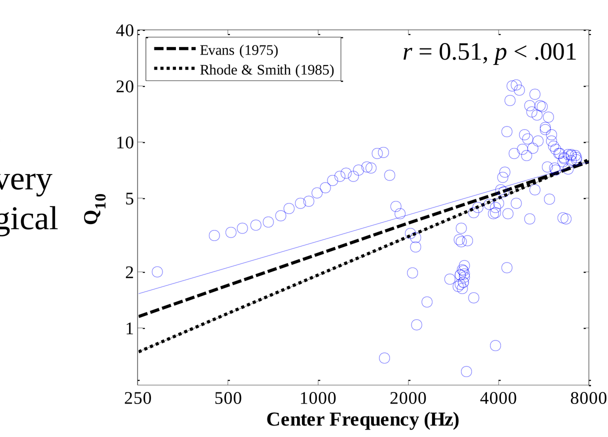
Tahitian (3 talkers)
Regions: Polynesian Triangle, Tahiti
Family: Polynesian Languages, Austronesian

Excellent fit to physiological measures



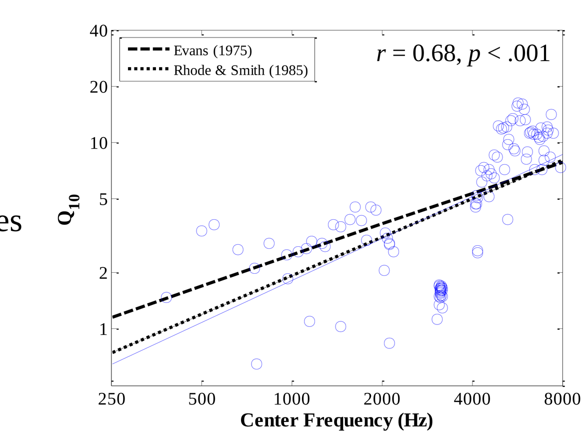
Urhobo (5 talkers)
Regions: Nigeria
Family: Niger-Congo

Filters are more variable, but still a very good fit to physiological measures



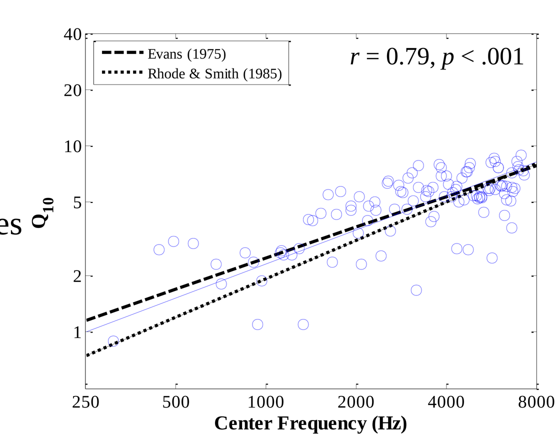
Vietnamese (1 talker)
Regions: Vietnam, Parts of Kampuchea (Cambodia), Thailand, Laos and oversea communities
Family: Muong-Vietnamese subgroup of the Mon-Khmer subfamily of the Austro-Asiatic family

Excellent fit to physiological measures



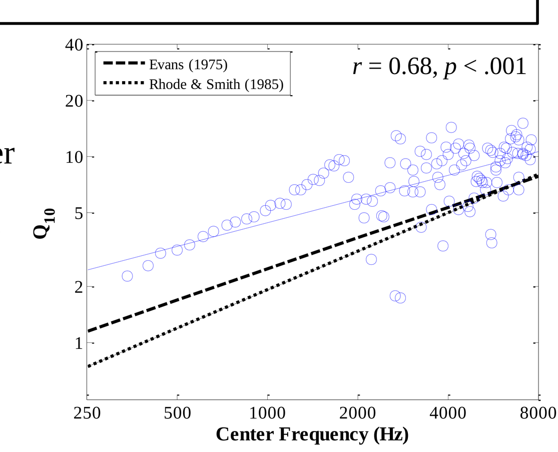
Wari (3 talkers)
Regions: Brazil
Family: Chupacura, Madeira

Excellent fit to physiological measures



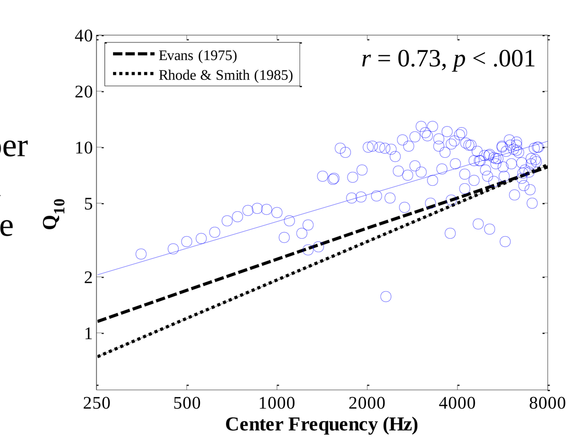
Xhosa (11 talkers)
Regions: South West Cape Province and Transkei in the Republic of South Africa
Family: Nguni group of the Bantu sub branch of the Benue-Congo brand of the Niger-Congo subfamily of the Niger-Khordofanian family

Filters slightly sharper than physiological measures; otherwise excellent fit



Yeyi (5 talkers)
Regions: Northwest Botswana, Namibia, East Caprivi, Ngamilan
Family: Bantu

Filters slightly sharper than physiological measures; otherwise excellent fit



DISCUSSION

Filters that optimally encode speech sounds in a wide variety of languages generally align with tuning properties in the mammalian auditory nerve. In most cases (especially cases where multiple talkers are sampled), slightly sharper filters are needed to encode speech than what is measured in the cat auditory system (evident in higher y-intercepts and/or steeper slopes for regression fits to speech). This agrees with recent data suggesting humans have sharper cochlear tuning than cats and other laboratory animals (Shera *et al.*, 2002; Joris *et al.*, 2011). In all, 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.