VOICEBOX: Speech Processing Toolbox for MATLAB

Introduction

VOICEBOX is a speech processing toolbox consists of MATLAB routines that are maintained by and mostly written by Mike Brookes, Department of Electrical & Electronic Engineering, Imperial College, Exhibition Road, London SW7 2BT, UK. Several of the routines require MATLAB V6.5 or above and require (normally slight) modification to work with earlier veresions.

The routines are available as a zip archive and are made available under the terms of the GNU Public License.

The routine VOICEBOX.M contains various installation-dependent parameters which may need to be altered before using the toolbox. In particular it contains a number of default directory paths indicating where temporary files should be created, where speech data normally resides, etc. You can override these defaults by editing voicebox.m directly or, more conveniently, by setting an environment variable VOICEBOX to the path of an initializing m-file. See the comments in voicebox.m for a fuller description.

For reading compressed SPHERE format files, you will need the SHORTEN program written by Tony Robinson and SoftSound Limited www.softsound.com. The path to the shorten executable must be set in voicebox.m.Unfortunately, the current version does not work on 64-bit systems.

MATLAB doesn't really like unicode fonts; some non-unicode fonts containing IPA phonetic symbols developed by SIL are available here.

Please send any comments, suggestions, bug reports etc to mike.brookes@ic.ac.uk.

Contents

Linear Predictive Coding routines

Speech Synthesis

Text-to-speech synthesis and glottal waveform models

Speech Enhancement

Spectral noise subtraction

Speech Coding

PCM coding, Vector quantisation

Speech Recognition

Front-end processing for recognition

Signal Processing

Miscellaneous signal processing functions

Information Theory

Routines for entropy calculation and symbol codes

Computer Vision

Routines for 3D rotation

Printing and Display Functions

Utilities for printing and graphics

Voicebox Parameters and System Interface

Get or set VOICEBOX and WINDOWS system parameters

Utility Functions

Miscellaneous utility functions

Audio File Input/Output

Routines are available to read and, in some cases write, a variety of file formats:

Read	Write	Suffix	
readwav	writewav	.wav	These routines allow an arbitrary number of channels and can deal with linear PCM (any precision up to 32 bits), A-law PCM, Mu-law PCM and Floating point formats. Large files can be read and written in small chunks.
readhtk	writehtk	.htk	Read and write waveform and parameter files used by Microsoft's Hidden Markov Toolkit.
readsfs		.sfs	Speech Filing system files from Mark Huckvale at UCL.
readsph		.sph	NIST Sphere format files (including TIMIT). Needs SHORTEN for compressed files.
readaif		.aif	AIFF format (Audio Interchange File Format) used by Mac users.
readcnx		cnx	Read Connex database files (from BT)
readau		au	Read AV audio files (from Sun)

Frequency Scale Conversion

From *f*	**To *f***	**Scale**	
frq2mel	mel2frq	mel	The *mel scale* is based on the human perception of sinewave pitch.
frq2erb	erb2frq	erb	The *erb* scale is based on the equivalent rectangular bandwidths of the human ear.
frq2bark	bark2frq	bark	The *bark* scale is based on critical bands and masking in the human ear.
frq2midi	midi2frq	midi	The *midi standard* specifies a numbering of *semitones* with middle C being 60. They can use the normal equal tempered scale or else the pythagorean scale of just intonation. They will in addition output note names in a character format.

Fourier, DCT and Hartley Transforms

Forward	**Inverse**	
rfft	irfft	Forward and inverse discrete fourier transforms on real data. Only the first half of the conjugate symmetric transform is generated. For even length data, the inverse routine is asumptotically twice as fast as the built-in MATLAB routine.
rsfft		Forward transform of real, symmetric data to give the first half only of the real, symmetric transform.
zoomfft		Calculate the discrete fourier transform at an arbitrary set of linearly spaced frequencies. Can be used to zoom into a subset of the full frequency range.
rdct	irdct	Forward and inverse discrete cosine transform on real data.
rhartley	rhartley	Hartley transform on real data (forward and inverse transforms are the same).

Random Numbers and Probability Distributions

- **Random Number Generation**

randvec	generates random vectors from gaussian or lognormal mixture distributions.
randiscr	generates discrete random values with a specified probability vector
stdspectrum	generates noise samples or filter coefficients for a variety of standard spectra including: A, B, C or BS468 weighting, USASI noise, POTS spectrum, LTASS, Internal masking noise (from SII spec)
randfilt	generates filtered gaussian noise without any startup transients.
rnsubset	selects a random subset of *k* elements from the numbers 1:*n*

- **Probability Density Functions**

lognmpdf	calculates the pdf of a lognormal distribution
gaussmix	generates a multivariate Gaussian mixture model (GMM) from training data
gaussmixg	calculates the global mean, covariance matrix and mode of a GMM
gaussmixp	calculates full and marginal log probability and relative mixture probabilities from a GMM
gaussmixd	determines marginal and conditional distributions from a GMM and can be used to perform inference on unobserved variables.

- **Miscellaneous**

histndim	calculates an n-dimensional histogram (and plots a 2-D one)
gausprod	calculates the product of two gaussian distributions
maxgauss	calculates the mean and variance of the maximum element of a gaussian vector

Vector Distance

disteusq	calculates the squared euclidean distance between all pairs of rows of two matrices.
distitar	calculates the Itakura spectral distances between sets of AR coefficients.
distitpf	calculates the Itakura spectral distances between power spectra.
distisar	calculates the Itakura-Saito spectral distances between sets of AR coefficients.
distispf	calculates the Itakura-Saito spectral distances between power spectra.
distchar	calculates the COSH spectral distances between sets of AR coefficients.
distchpf	calculates the COSH spectral distances between power spectra.

Speech Analysis

activlev	calculates the active level of a speech segment according to ITU-T recommendation P.56.
dypsa	estimates the glottal closure instants from the speech waveform.
enframe	can be used to split a signal up into frames. It can optionally apply a window to each frame.
correlogram	Calculates a 3D correlogram [slowly]
ewgrpdel	calculates the energy-weighted group delay waveform.
fram2wav	interpolates a sequence of frame-based value into a waveform
fxpefac	PEFAC pitch tracker
fxrapt	is an implementation of the RAPT pitch tracker by David Talkin.

gammabank	Determine a bank of IIR gammatone filters
importsii	calculate the SII importance function
mos2pesq	Convert MOS values to PESQ speech quality scores
overlapadd	Join frames up using overlap-add processing. Commonly used with enframe.
pesq2mos	Convert PESQ speech quality scores to MOS values
psycdigit	experimental estimation of monotonic/unimodal psychometric function using TIDIGITS
psycest	experimental estimation of monotonic psychometric function
psycestu	experimental estimation of unimodal psychometric function
psychofunc	calculate psychometric function
sigma	estimate glottal opening and closure instants from the laryngograph/EGG waveform
snrseg	calculate segmental SNR and global SNR relative to a reference signal
soundspeed	gives the speed of sound as a function of temperature
spgrambw	draws a spectrogram with many options. See tutorial.
txalign	finds the best alignment (in a least squares sense) between two sets of time markers (e.g. glottal closure instants).
vadsohn	voice activity detector

LPC Analysis of Speech

lpcauto & lpccovar	perform linear predictive coding (LPC) analysis. The routines relating to LPC are described in more detail on another page. A large number of conversion routines are included for changing the form of the LPC coefficients (e.g. AR coefficients, reflection coefficients etc.): these are of the form lpcxx2yy where xx and yy denote the coefficient sets.
lpcrr2am	calculates LPC filters for all orders up to a given maximum.
lpcbwexp	performs bandwidth expansion on an LPC filter.
ccwarpf	performs frequency warping in the complex cepstrum domain.
lpcifilt	performs inverse filtering to estimate the glottal waveform from the speech signal and the lpc coefficients.
lpcrand	can be used to generate random, stable filters for testing purposes.

Speech Synthesis

sapisynth	Text-to-speech synthesis (TTS) of a string or matrix entries
glotros	Calculates the Rosenberg model of the glottal flow waveform

glotlf	Calculates the Liljencrants-Fant model of the glottal flow waveform

Speech Enhancement

estnoiseg	uses an MMSE algorithm to estimate the noise spectrum from a noisy speech signal that has been divided into frames.
estnoisem	uses a minimum-statistics algorithm to estimate the noise spectrum from a noisy speech signal that has been divided into frames.
specsub	performs speech enhancement using spectral subtraction
ssubmmse	performs speech enhancement using the MMSE or log MMSE criteria

Speech Coding

lin2pcma	converts an audio waveform to 8-bit A-law PCM format
lin2pcmu	converts an audio waveform to 8-bit mu-law PCM format
pcma2lin	converts 8-bit A-law PCM to a waveform
pcmu2lin	converts 8-bit mu-law PCM to a waveform
kmeans	vector quantisation using the K-means algorithm
kmeanlbg	vector quantisation using the LBG algorithm
kmeanhar	vector quantisation using the K-harmonic means algorithm
potsband	calculates a bandpass filter corresponding to the standard telephone passband.

Speech Recognition

melcepst	implements a mel-cepstrum front end for a recogniser
melbankm	constructs a bandpass filterbank with mel-spaced centre frequencies
cep2pow	converts multivariate Gaussian means and covariances from the log power or cepstral domain to the power domain
pow2cep	converts multivariate Gaussian means and covariances from the power domain to the log power or cepstral domain
ldatrace	performs Linear Discriminant Analysis with optional constraints on the transform matrix

Signal Processing

ditherq	adds dither and quantizes a signal
dlyapsq	solves the discrete lyapunov equation using an efficient square root algorithm
findpeaks	finds the peaks in a signal
filterbank	Apply a bank of IIR filters to a signal
maxfilt	performs running maximum filter
meansqtf	calculates the output power of a rational filter with a white noise input
momfilt	generate running moments from a signal
sigalign	align a clean reference with a noise signal and find optimum gain
schmitt	passes a signal through a schmitt trigger having hysteresis
teager	calculate the Teager energy waveform
windows	generates window functions
windinfo	calculate window properties and figures of merit
zerocros	finds the zero crossings of a signal with interpolation

Information Theory

huffman	calculates optimum D-ary symbol code from a probability mass vector
entropy	calculates entropy and conditional entropy for discrete and continuous distributions

Computer Vision

imagehomog	Apply a homography transformation to an image with bilinear interpolation
polygonarea	Calculates the area of a polygon
polygonwind	Determines whether points are inside or outside a polygon
polygonxline	Determines where a line crosses a polygon
qrabs	Absolute value of a real quaternion
qrdivide	divide two real quaternions (or invert one)
qrdotdiv	elmentwise division of two real quaternion arrays
qrdotmult	elmentwise multiplication of two real quaternion arrays
qrmult	multiply two real quaternion arrays
qrpermute	permute the indices of a quaternion array
rectifyhomog	Apply rectifing homographies to a set of cameras to make their optical axes parallel
rot--2--	converts between the following representations of rotations: rotation matrix (ro), euler angles (eu), axis of rotation (ax), plane of rotation (pl), real quaternion vector (qr), real quaternion matrix (mr), complex quaternion vector (qc), complex quaternion matrix (mc). A detailed

	description is given here.
rotqrmean	Find the average of several rotation quaternions
rotqrvec	Apply a quaternion rotation to an array of 3D vectors
skew3d	Convert between vectors and skew symmetric matrices: 3x3 matrix <-> 3x1 vector and 4x4 Plucker matrix <-> 6x1 vector.
sphrharm	forward and inverse spherical harmonic transform using uniform, Gaussian or arbitrary inclination (elevation) grids and a uniform azimuth grid.
upolyhedron	Calculate the vertex coordinates and other characteristics of a uniform polyhedron

Printing and Display Functions

axisenlarge	enlarge the axes of a figure slightly
figbolden	makes the lines on a figure bold and enlarges font sizes for printing clearly
fig2emf	optionally makes the lines on a figure bold and then saves in windows metafile format
xticksi	Label the x-axis tick marks using SI multipliers for large and small values. Particularly useful for logarithmic plots.
yticksi	Label the y-axis tick marks using SI multipliers for large and small values. Particularly useful for logarithmic plots.
sprintsi	prints a value with the correct standard SI multiplier (e.g. 2100 prints as 2.1 k)
bitsprec	rounds values to a precision of n bits
frac2bin	converts numbers to fixed-point binary strings
v_colormap	set and display colormap information including colormaps that print well in monochrome

Voicebox Parameters and System Interface

voicebox	contains a number of installation-dependent global parameters and is likely to need editing for each particular setup.
unixwhich	searches the WINDOWS system path for an executable (like UNIX which command)
winenvar	Obtains WINDOWS environment variables

Utility Functions

atan2sc	arctangent function that returns the sin and cos of the angle
bitsprec	Rounds values to a precision of n bits
choosenk	all possible ways of choosing k elements out of the numbers 1:n without duplications
choosrnk	all possible ways of choosing k elements out of the numbers 1:n with duplications

	allowed
dlyapsq	Solve the discrete lyapunov equation
dualdiag	simultaneously diagonalises two matrices: this is useful in computing LDA or IMELDA transforms.
finishat	Estimate the finishing time of a long loop
fopenmkd	Equivalent to FOPEN() but creates any missing directories/folders
hostipinfo	Gives information about computer name and internet connections
logsum	calculates log(sum(exp(x))) without overflow problems.
minspane	Calculates the minimum spanning tree (a.k.a. shortest spanning tree) of a set of n-dimensional points
mintrace	Find a row permutation to minimize the trace of a matrix
m2htmlpwd	Create HTML documentation of matlab routines in the current directory
nearnonz	Replace zero elements by the nearest non-zero elements
permutes	all possible permutations of the numbers 1:n
quadpeak	find a quadratically-interpolated peak in a N-dimensional array by fitting a quadratic function to the array values
rotation	generates rotation matrices
zerotrim	removes from a matrix any trailing rows and columns that are all zero.