

Package ‘soundgen’

September 4, 2017

Type Package

Title Parametric Voice Synthesis

Version 1.0.0

Maintainer Andrey Anikin <rty.anik@rambler.ru>

URL <http://cogsci.se/soundgen.html>

Description Tools for sound synthesis and acoustic analysis.

Performs parametric synthesis of sounds with harmonic and noise components such as animal vocalizations or human voice. Also includes tools for spectral analysis, pitch tracking, audio segmentation, self-similarity matrices, and morphing.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports stats,
graphics,
utils,
tuneR,
seewave,
zoo,
shiny,
shinyBS,
reshape2,
mvtnorm,
plyr,
dtw,
grid,
phonTools

Depends R (>= 3.4)

RoxygenNote 6.0.1

Suggests knitr,
rmarkdown

VignetteBuilder knitr

R topics documented:

addVectors	2
analyze	3
analyzeFolder	7
compareSounds	10
crossFade	12
defaults	13
getRolloff	14
getSigmoid	15
getSmoothContour	16
getSpectralEnvelope	18
HzToSemitones	20
matchPars	21
morph	23
notesDict	24
optimizePars	24
permittedValues	26
pitchManual	27
playme	28
presets	28
segment	29
segmentFolder	30
segmentManual	32
semitonesToHz	33
soundgen	33
soundgen_app	37
spectrogram	38
ssm	40
Index	42

addVectors	<i>Add overlapping vectors</i>
------------	--------------------------------

Description

Adds two partly overlapping vectors, such as two waveforms, to produce a longer vector. The location at which vector 2 is pasted is defined by insertionPoint. Algorithm: both vectors are padded with zeros to match in length and then added. All NA's are converted to 0.

Usage

```
addVectors(v1, v2, insertionPoint)
```

Arguments

`v1, v2` numeric vectors

`insertionPoint` the index of element in vector 1 at which vector 2 will be inserted (any integer, can also be negative)

Examples

```
v1 = 1:6
v2 = rep(100, 3)
addVectors(v1, v2, insertionPoint = 5)
addVectors(v1, v2, insertionPoint = -4)
# note the asymmetry: insertionPoint refers to the first arg
addVectors(v2, v1, insertionPoint = -4)

v3 = rep(100, 15)
addVectors(v1, v3, insertionPoint = -4)
addVectors(v2, v3, insertionPoint = 7)
```

analyze

Analyze sound

Description

Acoustic analysis of a single sound file: pitch tracking and basic spectral characteristics. The default values of arguments are optimized for human non-linguistic vocalizations. See the vignette on acoustic analysis for details.

Usage

```
analyze(x, samplingRate = NULL, silence = 0.04, windowLength = 50,
  step = NULL, overlap = 50, wn = "gaussian", zp = 0, cutFreq = 6000,
  nFormants = 3, pitchMethods = c("autocor", "spec", "dom"),
  entropyThres = 0.6, pitchFloor = 75, pitchCeiling = 3500,
  priorMean = HzToSemitones(300), priorSD = 6, priorPlot = FALSE,
  nCands = 1, minVoicedCands = "autom", domThres = 0.1, domSmooth = 220,
  autocorThres = 0.7, autocorSmooth = NULL, cepThres = 0.3,
  cepSmooth = NULL, cepZp = 0, specThres = 0.3, specPeak = 0.35,
  specSinglePeakCert = 0.4, specHNRSlope = 0.8, specSmooth = 150,
  specMerge = 1, shortestSyl = 20, shortestPause = 60, interpolWin = 3,
  interpolTol = 0.3, interpolCert = 0.3, pathfinding = c("none", "fast",
  "slow")[2], annealPars = list(maxit = 5000, temp = 1000),
  certWeight = 0.5, snakeStep = 0.05, snakePlot = FALSE, smooth = 1,
  smoothVars = c("pitch", "dom"), summary = FALSE, plot = TRUE,
  savePath = NA, specPlot = list(contrast = 0.2, brightness = 0, ylim = c(0,
  5)), pitchPlot = list(col = rgb(0, 0, 1, 0.75), lwd = 3),
  candPlot = list(levels = c("autocor", "spec", "dom", "cep"), col =
  c("green", "red", "orange", "violet"), pch = c(16, 2, 3, 7), cex = 2))
```

Arguments

<code>x</code>	path to a .wav file or a vector of amplitudes with specified <code>samplingRate</code>
<code>samplingRate</code>	sampling rate of <code>x</code> (only needed if <code>x</code> is a numeric vector, rather than a .wav file)
<code>silence</code>	(0 to 1) frames with mean abs amplitude below <code>silence</code> threshold are not analyzed at all. NB: this number is dynamically updated: the actual <code>silence</code> threshold may be higher depending on the quietest frame, but it will never be lower than this specified number.
<code>windowLength</code>	length of FFT window, ms
<code>step</code>	you can override <code>overlap</code> by specifying FFT step, ms
<code>overlap</code>	overlap between successive FFT frames, %
<code>wn</code>	window type: gaussian, hanning, hamming, bartlett, rectangular, blackman, flat-top
<code>zp</code>	window length after zero padding, points
<code>cutFreq</code>	(>0 to Nyquist, Hz) repeat the calculation of spectral descriptives after discarding all info above <code>cutFreq</code> . Recommended if the original sampling rate varies across different analyzed audio files
<code>nFormants</code>	the number of formants to extract per FFT frame. Calls <code>findformants</code> with default settings
<code>pitchMethods</code>	methods of pitch estimation to consider for determining pitch contour: 'autocor' = autocorrelation (~PRAAT), 'cep' = cepstral, 'spec' = spectral (~BaNa), 'dom' = lowest dominant frequency band
<code>entropyThres</code>	pitch tracking is not performed for frames with Weiner entropy above <code>entropyThres</code> , but other spectral descriptives are still calculated
<code>pitchFloor, pitchCeiling</code>	absolute bounds for pitch candidates (Hz)
<code>priorMean, priorSD</code>	specifies the mean and sd of gamma distribution describing our prior knowledge about the most likely pitch values for this file. Specified in semitones: <code>priorMean = HzToSemitones(300), priorSD = 6</code> gives a prior with mean = 300 Hz and SD of 6 semitones (half an octave)
<code>priorPlot</code>	if TRUE, produces a separate plot of the prior
<code>nCands</code>	maximum number of pitch candidates per method (except for <code>dom</code> , which returns at most one candidate per frame), normally 1...4
<code>minVoicedCands</code>	minimum number of pitch candidates that have to be defined to consider a frame voiced (defaults to 2 if <code>dom</code> is among other candidates and 1 otherwise)
<code>domThres</code>	(0 to 1) to find the lowest dominant frequency band, we do short-term FFT and take the lowest frequency with amplitude at least <code>domThres</code>
<code>domSmooth</code>	the width of smoothing interval (Hz) for finding <code>dom</code>
<code>autocorThres, cepThres, specThres</code>	(0 to 1) separate voicing thresholds for detecting pitch candidates with three different methods: autocorrelation, cepstrum, and BaNa algorithm (see Details). Note that HNR is calculated even for unvoiced frames.

autocorSmooth	the width of smoothing interval (in bins) for finding peaks in the autocorrelation function. Defaults to 7 for sampling rate 44100 and smaller odd numbers for lower values of sampling rate
cepSmooth	the width of smoothing interval (in bins) for finding peaks in the cepstrum. Defaults to 31 for sampling rate 44100 and smaller odd numbers for lower values of sampling rate
cepZp	zero-padding of the spectrum used for cepstral pitch detection (final length of spectrum after zero-padding in points, e.g. 2^{13})
specPeak, specHNRSlope	when looking for putative harmonics in the spectrum, the threshold for peak detection is calculated as $\text{specPeak} * (1 - \text{HNR} * \text{specHNRSlope})$
specSinglePeakCert	(0 to 1) if F0 is calculated based on a single harmonic ratio (as opposed to several ratios converging on the same candidate), its certainty is taken to be specSinglePeakCert
specSmooth	the width of window for detecting peaks in the spectrum, Hz
specMerge	pitch candidates within specMerge semitones are merged with boosted certainty
shortestSyl	the smallest length of a voiced segment (ms) that constitutes a voiced syllable (shorter segments will be replaced by NA, as if unvoiced)
shortestPause	the smallest gap between voiced syllables (ms) that means they shouldn't be merged into one voiced syllable
interpWin, interpTol, interpCert	control the behavior of interpolation algorithm when postprocessing pitch candidates. To turn off interpolation, set interpWin to NULL. See <code>soundgen:::pathfinder</code> for details.
pathfinding	method of finding the optimal path through pitch candidates: 'none' = best candidate per frame, 'fast' = simple heuristic, 'slow' = annealing. See <code>soundgen:::pathfinder</code>
annealPars	a list of control parameters for postprocessing of pitch contour with SANN algorithm of <code>optim</code> . This is only relevant if <code>pathfinding = 'slow'</code>
certWeight	(0 to 1) in pitch postprocessing, specifies how much we prioritize the certainty of pitch candidates vs. pitch jumps / the internal tension of the resulting pitch curve
snakeStep	optimized path through pitch candidates is further processed to minimize the elastic force acting on pitch contour. To disable, set snakeStep to NULL
snakePlot	if TRUE, plots the snake
smooth, smoothVars	if smooth is a positive number, outliers of the variables in smoothVars are adjusted with median smoothing. smooth of 1 corresponds to a window of ~100 ms and tolerated deviation of ~4 semitones. To disable, set smooth to NULL
summary	if TRUE, returns only a summary of the measured acoustic variables (mean, median and SD). If FALSE, returns a list containing frame-by-frame values
plot	if TRUE, produces a spectrogram with pitch contour overlaid
savePath	if a valid path is specified, a plot is saved in this folder (defaults to NA)

specPlot	a list of graphical parameters passed to <code>spectrogram</code> . Set to NULL to suppress plotting just the spectrogram
pitchPlot	a list of graphical parameters for displaying the final pitch contour. Set to NULL or NA to suppress
candPlot	a list of graphical parameters for displaying individual pitch candidates. Set to NULL or NA to suppress

Value

If `summary = TRUE`, returns a dataframe with one row and three column per acoustic variable (mean / median / SD). If `summary = FALSE`, returns a dataframe with one row per FFT frame and one column per acoustic variable. The best guess at the pitch contour considering all available information is stored in the variable called "pitch". In addition, the output contains a number of other acoustic descriptors and pitch estimates by separate algorithms included in `pitchMethods`. See the vignette on acoustic analysis for a full explanation of returned measures.

Examples

```
## Not run:
sound1 = soundgen(syllLen = 900, pitchAnchors = list(
  time = c(0, .3, .8, 1), value = c(300, 900, 400, 2300)),
  noiseAnchors = list(time = c(0, 900), value = c(-40, 00)),
  temperature = 0)
# playme(sound1, 16000)
a1 = analyze(sound1, samplingRate = 16000, plot = TRUE)
# or, to improve the quality of postprocessing:
a1 = analyze(sound1, samplingRate = 16000, plot = TRUE, pathfinding = 'slow')
median(a1$pitch, na.rm = TRUE) # 614 Hz
# (can vary, since postprocessing is stochastic)
# compare to the true value:
median(getSmoothContour(anchors = list(time = c(0, .3, .8, 1),
  value = c(300, 900, 400, 2300)), len = 1000))

# the same pitch contour, but harder b/c of subharmonics and jitter
sound2 = soundgen(syllLen = 900, temperature = 0,
  pitchAnchors = list(time = c(0, .3, .8, 1),
    value = c(300, 900, 400, 2300)),
  noiseAnchors = list(time = c(0, 900), value = c(-40, 20)),
  subDep = 100, jitterDep = 0.5, pitchEffectsAmount = 100)
sound2 = soundgen(syllLen = 900, pitchAnchors = list(
  time = c(0, .3, .8, 1), value = c(300, 900, 400, 2300)),
  noiseAnchors = list(time = c(0, 900), value = c(-40, 20)),
  subDep = 100, jitterDep = 0.5, pitchEffects_amount = 100, temperature = 0)
# playme(sound2, 16000)
a2 = analyze(sound2, samplingRate = 16000, plot = TRUE, pathfinding = 'slow')
# many candidates are off, but the overall contour should be mostly accurate

# Fancy plotting options:
a = analyze(sound2, samplingRate = 16000, plot = TRUE,
  specPlot = list(xlab = 'Time, ms', colorTheme = 'seewave', contrast = .8),
  candPlot = list(cex = 3, col = c('gray70', 'yellow', 'purple', 'maroon')))
```

```

pitchPlot = list(col = 'black', lty = 3, lwd = 3))

# Plot pitch candidates w/o a spectrogram
a = analyze(sound2, samplingRate = 16000, plot = TRUE, specPlot = NA)

## End(Not run)

```

analyzeFolder

Analyze sound

Description

Acoustic analysis of all .wav files in a folder.

Usage

```

analyzeFolder(myfolder, verbose = TRUE, samplingRate = NULL,
  silence = 0.04, windowLength = 50, step = NULL, overlap = 50,
  wn = "gaussian", zp = 0, cutFreq = 6000, nFormants = 3,
  pitchMethods = c("autocor", "spec", "dom"), entropyThres = 0.6,
  pitchFloor = 75, pitchCeiling = 3500, priorMean = HzToSemitones(300),
  priorSD = 6, priorPlot = FALSE, nCands = 1, minVoicedCands = "autom",
  domThres = 0.1, domSmooth = 220, autocorThres = 0.7,
  autocorSmooth = NULL, cepThres = 0.3, cepSmooth = NULL, cepZp = 0,
  specThres = 0.3, specPeak = 0.35, specSinglePeakCert = 0.4,
  specHNRSlope = 0.8, specSmooth = 150, specMerge = 1, shortestSyl = 20,
  shortestPause = 60, interpolWin = 3, interpolTol = 0.3,
  interpolCert = 0.3, pathfinding = c("none", "fast", "slow")[2],
  annealPars = list(maxit = 5000, temp = 1000), certWeight = 0.5,
  snakeStep = 0.05, snakePlot = FALSE, smooth = 1,
  smoothVars = c("pitch", "dom"), summary = TRUE, plot = FALSE,
  savePath = NA, specPlot = list(contrast = 0.2, brightness = 0, ylim = c(0,
  5)), pitchPlot = list(col = rgb(0, 0, 1, 0.75), lwd = 3),
  candPlot = list(levels = c("autocor", "spec", "dom", "cep"), col =
  c("green", "red", "orange", "violet"), pch = c(16, 2, 3, 7), cex = 2))

```

Arguments

myfolder	full path to target folder
verbose	if TRUE, reports progress and estimated time left
samplingRate	sampling rate of x (only needed if x is a numeric vector, rather than a .wav file)
silence	(0 to 1) frames with mean abs amplitude below silence threshold are not analyzed at all. NB: this number is dynamically updated: the actual silence threshold may be higher depending on the quietest frame, but it will never be lower than this specified number.
windowLength	length of FFT window, ms

step	you can override overlap by specifying FFT step, ms
overlap	overlap between successive FFT frames, %
wn	window type: gaussian, hanning, hamming, bartlett, rectangular, blackman, flat-top
zp	window length after zero padding, points
cutFreq	(>0 to Nyquist, Hz) repeat the calculation of spectral descriptives after discarding all info above cutFreq. Recommended if the original sampling rate varies across different analyzed audio files
nFormants	the number of formants to extract per FFT frame. Calls <code>findformants</code> with default settings
pitchMethods	methods of pitch estimation to consider for determining pitch contour: 'autocor' = autocorrelation (~PRAAT), 'cep' = cepstral, 'spec' = spectral (~BaNa), 'dom' = lowest dominant frequency band
entropyThres	pitch tracking is not performed for frames with Weiner entropy above entropyThres, but other spectral descriptives are still calculated
pitchFloor	absolute bounds for pitch candidates (Hz)
pitchCeiling	absolute bounds for pitch candidates (Hz)
priorMean	specifies the mean and sd of gamma distribution describing our prior knowledge about the most likely pitch values for this file. Specified in semitones: <code>priorMean = HzToSemitones(300), priorSD = 6</code> gives a prior with mean = 300 Hz and SD of 6 semitones (half an octave)
priorSD	specifies the mean and sd of gamma distribution describing our prior knowledge about the most likely pitch values for this file. Specified in semitones: <code>priorMean = HzToSemitones(300), priorSD = 6</code> gives a prior with mean = 300 Hz and SD of 6 semitones (half an octave)
priorPlot	if TRUE, produces a separate plot of the prior
nCands	maximum number of pitch candidates per method (except for dom, which returns at most one candidate per frame), normally 1..4
minVoicedCands	minimum number of pitch candidates that have to be defined to consider a frame voiced (defaults to 2 if dom is among other candidates and 1 otherwise)
domThres	(0 to 1) to find the lowest dominant frequency band, we do short-term FFT and take the lowest frequency with amplitude at least domThres
domSmooth	the width of smoothing interval (Hz) for finding dom
autocorThres	(0 to 1) separate voicing thresholds for detecting pitch candidates with three different methods: autocorrelation, cepstrum, and BaNa algorithm (see Details). Note that HNR is calculated even for unvoiced frames.
autocorSmooth	the width of smoothing interval (in bins) for finding peaks in the autocorrelation function. Defaults to 7 for sampling rate 44100 and smaller odd numbers for lower values of sampling rate
cepThres	(0 to 1) separate voicing thresholds for detecting pitch candidates with three different methods: autocorrelation, cepstrum, and BaNa algorithm (see Details). Note that HNR is calculated even for unvoiced frames.

cepSmooth	the width of smoothing interval (in bins) for finding peaks in the cepstrum. Defaults to 31 for sampling rate 44100 and smaller odd numbers for lower values of sampling rate
cepZp	zero-padding of the spectrum used for cepstral pitch detection (final length of spectrum after zero-padding in points, e.g. 2^{13})
specThres	(0 to 1) separate voicing thresholds for detecting pitch candidates with three different methods: autocorrelation, cepstrum, and BaNa algorithm (see Details). Note that HNR is calculated even for unvoiced frames.
specPeak	when looking for putative harmonics in the spectrum, the threshold for peak detection is calculated as $\text{specPeak} * (1 - \text{HNR} * \text{specHNRSlope})$
specSinglePeakCert	(0 to 1) if F0 is calculated based on a single harmonic ratio (as opposed to several ratios converging on the same candidate), its certainty is taken to be specSinglePeakCert
specHNRSlope	when looking for putative harmonics in the spectrum, the threshold for peak detection is calculated as $\text{specPeak} * (1 - \text{HNR} * \text{specHNRSlope})$
specSmooth	the width of window for detecting peaks in the spectrum, Hz
specMerge	pitch candidates within specMerge semitones are merged with boosted certainty
shortestSyl	the smallest length of a voiced segment (ms) that constitutes a voiced syllable (shorter segments will be replaced by NA, as if unvoiced)
shortestPause	the smallest gap between voiced syllables (ms) that means they shouldn't be merged into one voiced syllable
interpWin	control the behavior of interpolation algorithm when postprocessing pitch candidates. To turn off interpolation, set interpWin to NULL. See <code>soundgen:::pathfinder</code> for details.
interpTol	control the behavior of interpolation algorithm when postprocessing pitch candidates. To turn off interpolation, set interpWin to NULL. See <code>soundgen:::pathfinder</code> for details.
interpCert	control the behavior of interpolation algorithm when postprocessing pitch candidates. To turn off interpolation, set interpWin to NULL. See <code>soundgen:::pathfinder</code> for details.
pathfinding	method of finding the optimal path through pitch candidates: 'none' = best candidate per frame, 'fast' = simple heuristic, 'slow' = annealing. See <code>soundgen:::pathfinder</code>
annealPars	a list of control parameters for postprocessing of pitch contour with SANN algorithm of <code>optim</code> . This is only relevant if <code>pathfinding = 'slow'</code>
certWeight	(0 to 1) in pitch postprocessing, specifies how much we prioritize the certainty of pitch candidates vs. pitch jumps / the internal tension of the resulting pitch curve
snakeStep	optimized path through pitch candidates is further processed to minimize the elastic force acting on pitch contour. To disable, set snakeStep to NULL
snakePlot	if TRUE, plots the snake
smooth	if smooth is a positive number, outliers of the variables in smoothVars are adjusted with median smoothing. smooth of 1 corresponds to a window of ~100 ms and tolerated deviation of ~4 semitones. To disable, set smooth to NULL

smoothVars	if smooth is a positive number, outliers of the variables in smoothVars are adjusted with median smoothing. smooth of 1 corresponds to a window of ~100 ms and tolerated deviation of ~4 semitones. To disable, set smooth to NULL
summary	if TRUE, returns only a summary of the measured acoustic variables (mean, median and SD). If FALSE, returns a list containing frame-by-frame values
plot	if TRUE, produces a spectrogram with pitch contour overlaid
savePath	if a valid path is specified, a plot is saved in this folder (defaults to NA)
specPlot	a list of graphical parameters passed to spectrogram . Set to NULL to suppress plotting just the spectrogram
pitchPlot	a list of graphical parameters for displaying the final pitch contour. Set to NULL or NA to suppress
candPlot	a list of graphical parameters for displaying individual pitch candidates. Set to NULL or NA to suppress

Value

If summary is TRUE, returns a dataframe with one row per audio file. If summary is FALSE, returns a list of detailed descriptives.

Examples

```
## Not run:
# download 260 sounds from Anikin & Persson (2017)
# http://cogsci.se/personal/results/
# 01_anikin-persson_2016_naturalistics-non-linguistic-vocalizations/260sounds_wav.zip
# unzip them into a folder, say '~/Downloads/temp'
myfolder = '~/Downloads/temp' # 260 .wav files live here
s = analyzeFolder(myfolder, verbose = TRUE) # ~ 15-30 minutes!

# Check accuracy: import manually verified pitch values (our "key")
key = pitchManual # a vector of 260 floats
trial = s$pitch_median
cor(key, trial, use = 'pairwise.complete.obs')
plot(log(key), log(trial))
abline(a=0, b=1, col='red')

## End(Not run)
```

compareSounds

Compare sounds (experimental)

Description

Computes similarity between two sounds based on correlating mel-transformed spectra (auditory spectra). Called by [matchPars](#).

Usage

```
compareSounds(target, targetSpec = NULL, cand, samplingRate = NULL,
  method = c("cor", "cosine", "pixel", "dtw")[1:4], windowLength = 40,
  overlap = 50, step = NULL, padWith = NA, penalizeLengthDif = TRUE,
  throwaway = -120, maxFreq = NULL, summary = TRUE)
```

Arguments

target	the sound we want to reproduce using soundgen: path to a .wav file or numeric vector
targetSpec	if already calculated, the target auditory spectrum can be provided to speed things up
cand	the sound to be compared to target
samplingRate	sampling rate of target (only needed if target is a numeric vector, rather than a .wav file)
method	method of comparing mel-transformed spectra of two sounds: "cor" = average Pearson's correlation of mel-transformed spectra of individual FFT frames; "cosine" = same as "cor" but with cosine similarity instead of Pearson's correlation; "pixel" = absolute difference between each point in the two spectra; "dtw" = discrete time warp with <code>dtw</code>
windowLength	length of FFT window, ms
overlap	overlap between successive FFT frames, %
step	you can override <code>overlap</code> by specifying FFT step, ms
padWith	compared spectra are padded with either silence (<code>padWith = 0</code>) or with NA's (<code>padWith = NA</code>) to have the same number of columns. When the sounds are of different duration, padding with zeros rather than NA's improves the fit to target measured by <code>method = 'pixel'</code> and <code>'dtw'</code> , but it has no effect on <code>'cor'</code> and <code>'cosine'</code> .
penalizeLengthDif	if TRUE, sounds of different length are considered to be less similar; if FALSE, only the overlapping parts of two sounds are compared
throwaway	parts of the spectra quieter than <code>throwaway</code> dB are not compared
maxFreq	parts of the spectra above <code>maxFreq</code> Hz are not compared
summary	if TRUE, returns the mean of similarity values calculated by all methods in <code>method</code>

Examples

```
## Not run:
target = soundgen(syllLen = 500, formants = 'a',
  pitchAnchors = data.frame(time = c(0, 0.1, 0.9, 1),
    value = c(100, 150, 135, 100)),
  temperature = 0)
targetSpec = soundgen::getMelSpec(target, samplingRate = 16000)
parsToTry = list(
  list(formants = 'i', # wrong
```

```

        pitchAnchors = data.frame(time = c(0, 1),           # wrong
                                   value = c(200, 300)),
list(formants = 'i',                                     # wrong
      pitchAnchors = data.frame(time = c(0, 0.1, 0.9, 1), # right
                                   value = c(100, 150, 135, 100))),
list(formants = 'a',                                     # right
      pitchAnchors = data.frame(time = c(0,1),          # wrong
                                   value = c(200, 300))),
list(formants = 'a',                                     # right
      pitchAnchors = data.frame(time = c(0, 0.1, 0.9, 1), # right
                                   value = c(100, 150, 135, 100)) # right
)

sounds = list()
for (s in 1:length(parsToTry)) {
  sounds[[length(sounds) + 1]] = do.call(soundgen,
    c(parsToTry[[s]], list(temperature = 0, syllLen = 500)))
}

method = c('cor', 'cosine', 'pixel', 'dtw')
df = matrix(NA, nrow = length(parsToTry), ncol = length(method))
colnames(df) = method
df = as.data.frame(df)
for (i in 1:nrow(df)) {
  df[i, ] = compareSounds(
    target = NULL, # can use target instead of targetSpec...
    targetSpec = targetSpec, # ...but faster to calculate targetSpec once
    cand = sounds[[i]],
    samplingRate = 16000,
    padWith = NA,
    penalizeLengthDif = TRUE,
    method = method,
    summary = FALSE
  )
}
df$av = rowMeans(df, na.rm = TRUE)
df # row 1 = wrong pitch & formants, ..., row 4 = right pitch & formants

## End(Not run)

```

crossFade

Join two waveforms by cross-fading

Description

crossFade joins two input vectors (waveforms) by cross-fading. It truncates both input vectors, so that amp1 ends with a zero crossing and amp2 starts with a zero crossing, both on an upward portion of the soundwave. Then it cross-fades both vectors linearly with an overlap of crossLen or crossLenPoints. If the input vectors are too short for the specified length of cross-faded region, the two vectors are concatenated at zero crossings instead of cross-fading. Soundgen uses crossFade

for gluing together epochs with different regimes of pitch effects (see the vignette on sound generation), but it can also be useful for joining two separately generated sounds without audible artifacts.

Usage

```
crossFade(amp11, amp12, samplingRate, crossLen = 15, crossLenPoints = NULL)
```

Arguments

`amp11`, `amp12` two numeric vectors (waveforms) to be joined

`samplingRate` the sampling rate of input vectors, Hz

`crossLen` the length of overlap, in ms (doesn't need to be specified if `crossLenPoints` is not NULL)

`crossLenPoints` (optional) the length of overlap, in points (defaults to NULL)

Value

Returns a numeric vector.

Examples

```
sound1 = sin(1:100 / 9)
sound2 = sin(7:107 / 3)
plot(c(sound1, sound2), type = 'b')
# an ugly discontinuity at 100 that will make an audible click

sound = crossFade(sound1, sound2, crossLenPoints = 5)
plot(sound, type = 'b') # a nice, smooth transition
length(sound) # but note that cross-fading costs us ~60 points
# because of trimming to zero crossings
```

defaults

Shiny app defaults

Description

A list of default values for Shiny app `soundgen_app()` - mostly the same as the defaults for `soundgen()`. NB: if defaults change, this has to be updated!!!

Usage

```
defaults
```

Format

An object of class `list` of length 51.

getRolloff *Control rolloff of harmonics*

Description

Harmonics are generated as separate sine waves. But we don't want each harmonic to be equally strong, so we normally specify some rolloff function that describes the loss of energy in upper harmonics relative to the fundamental frequency (f_0). `getRolloff` provides flexible control over this rolloff function, going beyond simple exponential decay (`rolloff`). Use quadratic terms to modify the behavior of a few lower harmonics, `rolloffOct` to adjust the rate of decay per octave, and `rolloffKHz` for rolloff correction depending on f_0 . Plot the output with different parameter values and see examples below and the vignette to get a feel for how to use `getRolloff` effectively.

Usage

```
getRolloff(pitch_per_gc = c(440), nHarmonics = 100, rolloff = -12,
           rolloffOct = -2, rolloffParab = 0, rolloffParabHarm = 2,
           rolloffParabCeiling = NULL, rolloffKHz = -6, baseline = 200,
           throwaway = -120, samplingRate = 16000, plot = FALSE)
```

Arguments

<code>pitch_per_gc</code>	a vector of f_0 per glottal cycle, Hz
<code>nHarmonics</code>	maximum number of harmonics to generate (very weak harmonics with amplitude < <code>throwaway</code> will be discarded)
<code>rolloff</code>	basic rolloff at a constant rate of <code>rolloff</code> db/octave (exponential decay). See <code>getRolloff</code> for more details
<code>rolloffOct</code>	basic rolloff changes from lower to upper harmonics (regardless of f_0) by <code>rolloffOct</code> dB/oct. For example, we can get steeper rolloff in the upper part of the spectrum
<code>rolloffParab</code>	an optional quadratic term affecting only the first <code>rolloffParabHarm</code> harmonics. The middle harmonic of the first <code>rolloffParabHarm</code> harmonics is amplified or dampened by <code>rolloffParab</code> dB relative to the basic exponential decay.
<code>rolloffParabHarm</code>	the number of harmonics affected by <code>rolloffParab</code>
<code>rolloffParabCeiling</code>	quadratic adjustment is applied only up to <code>rolloffParabCeiling</code> , Hz. If not NULL, it overrides <code>rolloffParabHarm</code>
<code>rolloffKHz</code>	rolloff changes linearly with f_0 by <code>rolloffKHz</code> dB/kHz. For ex., -6 dB/kHz gives a 6 dB steeper basic rolloff as f_0 goes up by 1000 Hz
<code>baseline</code>	The "neutral" frequency, at which no adjustment of rolloff takes place regardless of <code>rolloffKHz</code>
<code>throwaway</code>	discard harmonics and noise that are quieter than this number (in dB, defaults to -120) to save computational resources
<code>samplingRate</code>	sampling rate (needed to stop at Nyquist frequency and for plotting purposes)
<code>plot</code>	if TRUE, produces a plot

Value

Returns a matrix of amplitude multiplication factors for adjusting the amplitude of harmonics relative to f_0 . Each row of output contains one harmonic, and each column contains one glottal cycle.

Examples

```
# steady exponential rolloff of -12 dB per octave
rolloff = getRolloff(pitch_per_gc = 150, rolloff = -12,
  rolloffOct = 0, plot = TRUE)
# the rate of rolloff slows down with each octave
rolloff = getRolloff(pitch_per_gc = 150, rolloff = -12,
  rolloffOct = 2, plot = TRUE)
# the rate of rolloff increases with each octave
rolloff = getRolloff(pitch_per_gc = 150, rolloff = -12,
  rolloffOct = -2, plot = TRUE)

# variable f0: the lower f0, the more harmonics are non-zero
rolloff = getRolloff(pitch_per_gc = c(150, 800, 3000),
  rolloffOct = 0, plot = TRUE)
# without the correction for f0 (rolloffKHz),
# high-pitched sounds have the same rolloff as low-pitched sounds,
# producing unnaturally strong high-frequency harmonics
rolloff = getRolloff(pitch_per_gc = c(150, 800, 3000),
  rolloffOct = 0, rolloffKHz = 0, plot = TRUE)

# parabolic adjustment of lower harmonics
rolloff = getRolloff(pitch_per_gc = 350, rolloffParab = 0,
  rolloffParabHarm = 2, plot = TRUE)
# rolloffParabHarm = 1 affects only f0
rolloff = getRolloff(pitch_per_gc = 150, rolloffParab = 30,
  rolloffParabHarm = 1, plot = TRUE)
# rolloffParabHarm=2 or 3 affects only h1
rolloff = getRolloff(pitch_per_gc = 150, rolloffParab = 30,
  rolloffParabHarm = 2, plot = TRUE)
# rolloffParabHarm = 4 affects h1 and h2, etc
rolloff = getRolloff(pitch_per_gc = 150, rolloffParab = 30,
  rolloffParabHarm = 4, plot = TRUE)
# negative rolloffParab weakens lower harmonics
rolloff = getRolloff(pitch_per_gc = 150, rolloffParab = -20,
  rolloffParabHarm = 7, plot = TRUE)
# only harmonics below 2000 Hz are affected
rolloff = getRolloff(pitch_per_gc = c(150, 600),
  rolloffParab = -20, rolloffParabCeiling = 2000,
  plot = TRUE)
```

Description

Internal soundgen function.

Usage

```
getSigmoid(len, samplingRate = 16000, freq = 5, shape = 0,
           spikiness = 1)
```

Arguments

len	the length of output vector
samplingRate	the sampling rate of the output vector, Hz
freq	the frequency of amplitude modulation, Hz
shape	0 = ~sine, -1 = clicks, +1 = notches (NB: vice versa in soundgen!)
spikiness	the larger, the more quickly the shape of filter leaves sine-like approximation as shape deviates from 0

Details

Produces a filter for amplitude modulation ranging from clicks to approximately a sine wave to reversed clicks (small episodes of silence). The filter is made from concatenated sigmoids and their mirror reflections.

Value

Returns a vector of length len and range from 0 to 1

Examples

```
for (shape in c(0, -.1, .1, -1, 1)) {
  s = soundgen::getSigmoid(shape = shape, len = 1000, samplingRate = 500, freq = 2)
  plot(s, type = 'l', main = paste('shape =', shape), xlab = '', ylab = '')
}
```

getSmoothContour	<i>Smooth contour from anchors</i>
------------------	------------------------------------

Description

Returns a smooth contour based on an arbitrary number of anchors. Used by [soundgen](#) for generating intonation contour, mouth opening, etc. Note that pitch contours are treated as a special case: values are log-transformed prior to smoothing, so that with 2 anchors we get a linear transition on a log scale (as if we were operating with musical notes rather than frequencies in Hz). Pitch plots have two Y axes: one showing Hz and the other showing musical notation.

Usage

```
getSmoothContour(anchors = data.frame(time = c(0, 1), value = c(0, 1)),
  len = NULL, thisIsPitch = FALSE, method = c("spline", "loess")[2],
  valueFloor = NULL, valueCeiling = NULL, plot = FALSE, main = "",
  xlim = NULL, ylim = NULL, samplingRate = 16000, voiced = NULL,
  contourLabel = NULL, ...)
```

Arguments

anchors	a dataframe of anchors with specified time and amplitude. anchors\$time can be in ms (with len=NULL) or in arbitrary units, eg 0 to 1 (with duration determined by len, which must then be provided in ms). So anchors\$time is assumed to be in ms if len=NULL and relative if len is specified. anchors\$value can be on any scale.
len	the required length of the output contour. If NULL, it will be calculated based on the maximum time value (in ms) and samplingRate
thisIsPitch	(boolean) is this a pitch contour? If true, log-transforms before smoothing and plots in both Hz and musical notation
method	smoothing method: 'spline' or 'loess' (default). Consider 'spline' if you have a lot of anchors and not much patience, since it is much faster than 'loess', but it may produce weird results when there are only a few anchors
valueFloor, valueCeiling	lower/upper bounds for the contour
plot	(boolean) produce a plot?
main, xlim, ylim	plotting options
samplingRate	sampling rate used to convert time values to points (Hz)
voiced, contourLabel	graphical pars for plotting breathing contours (see examples below)
...	other plotting options passed to plot()

Value

Returns a numeric vector.

Examples

```
a = getSmoothContour(anchors = data.frame(
  time = c(50, 137, 300), value = c(0.03, 0.78, 0.5)),
  voiced = 200, valueFloor = 0, plot = TRUE, main = '',
  samplingRate = 16000) # breathing

a = getSmoothContour(anchors = data.frame(
  time = c(0, .1), value = c(350, 800)),
  len = 5500, thisIsPitch = TRUE, plot = TRUE,
  samplingRate = 3500) # pitch
```

```
# a single anchor gives constant pitch
a = getSmoothContour(anchors = data.frame(time = 0, value = 800),
  len = 500, thisIsPitch = TRUE, plot = TRUE, samplingRate = 500)

# two anchors give loglinear pitch change
a = getSmoothContour(anchors = data.frame(
  time = c(0, 1), value = c(220, 440)),
  len = 500, thisIsPitch = TRUE, plot = TRUE, samplingRate = 500)
```

getSpectralEnvelope *Spectral envelope*

Description

Prepares a spectral envelope for filtering a sound to add formants, lip radiation, and some stochastic component regulated by temperature. Formants are specified as a list containing time, frequency, amplitude, and width values for each formant (see examples). NB: each formant is generated as a gamma distribution with mean = freq and SD = width. Formant bandwidths in soundgen are therefore NOT compatible with formant bandwidths used in Klatt synthesizer and other algorithms that rely on FIR instead of FFT.

Usage

```
getSpectralEnvelope(nr, nc, formants = NA, formantDep = 1, rolloffLip = 6,
  mouthAnchors = NA, mouthOpenThres = 0, openMouthBoost = 0,
  vocalTract = NULL, temperature = 0, formDrift = 0.3, formDisp = 0.2,
  formantDepStoch = 30, smoothLinearFactor = 1, samplingRate = 16000,
  speedSound = 35400, plot = FALSE, duration = NULL,
  colorTheme = c("bw", "seewave", "...")[1], nCols = 100, xlab = "Time",
  ylab = "Frequency, kHz", ...)
```

Arguments

nr	the number of frequency bins = $\text{windowLength_points}/2$, where <code>windowLength_points</code> is the size of window for Fourier transform
nc	the number of time steps for Fourier transform
formants	either a character string like "aau" referring to default presets for speaker "M1" or a list of formant times, frequencies, amplitudes, and bandwidths. <code>formants = NA</code> defaults to schwa. Time stamps for formants and mouthOpening can be specified in ms or an any other arbitrary scale.
formantDep	scale factor of formant amplitude (1 = no change relative to amplitudes in formants)
rolloffLip	the effect of lip radiation on source spectrum, dB/oct (the default of +6 dB/oct produces a high-frequency boost when the mouth is open)
mouthAnchors	a numeric vector of mouth opening (0 to 1, 0.5 = neutral, i.e. no modification) or a dataframe specifying the time (ms) and value of mouth opening

mouthOpenThres	the mouth is considered to be open when its opening is greater than mouthOpenThres. Defaults to 0
openMouthBoost	amplify the voice when the mouth is open by openMouthBoost dB
vocalTract	the length of vocal tract, cm. Used for calculating formant dispersion (for adding extra formants) and formant transitions as the mouth opens and closes
temperature	hyperparameter for regulating the amount of stochasticity in sound generation
formDrift	scale factor regulating the effect of temperature on the depth of random drift of all formants (user-defined and stochastic): the higher, the more formants drift at a given temperature
formDisp	scale factor regulating the effect of temperature on the irregularity of the dispersion of stochastic formants: the higher, the more unevenly stochastic formants are spaced at a given temperature
formantDepStoch	the amplitude of additional formants added above the highest specified formant (only if temperature > 0)
smoothLinearFactor	regulates smoothing of formant anchors (0 to +Inf) as they are upsampled to the number of fft steps nc. This is necessary because the input formants normally contains fewer sets of formant values than the number of fft steps. smoothLinearFactor = 0: close to default spline; >3: approaches linear extrapolation
samplingRate	sampling frequency, Hz
speedSound	speed of sound in warm air, cm/s. Stevens (2000) "Acoustic phonetics", p. 138
plot	if TRUE, produces a plot of the spectral envelope
duration	duration of the sound, ms (for plotting purposes only)
colorTheme	black and white ('bw'), as in seewave package ('seewave'), or another color theme (e.g. 'heat.colors')
nCols	number of colors in the palette
xlab, ylab	labels of axes
...	other graphical parameters passed on to image()

Value

Returns a spectral filter (matrix nr x nc, where nr is the number of frequency bins = windowLength_points/2 and nc is the number of time steps)

Examples

```
# [a] with F1-F3 visible
image(t(getSpectralEnvelope(nr = 512, nc = 50,
  formants = soundgen::convertStringToFormants('a'),
  temperature = 0)))
# some "wiggling" of specified formants plus extra formants on top
image(t(getSpectralEnvelope(nr = 512, nc = 50,
  formants = soundgen::convertStringToFormants('a'),
  temperature = 0.1, formantDepStoch = 10)))
```

```

# stronger extra formants
image(t(getSpectralEnvelope(nr = 512, nc = 50,
  formants = soundgen::convertStringToFormants('a'),
  temperature = 0.1, formantDepStoch = 30)))
# a schwa based on the length of vocal tract = 15.5 cm
image(t(getSpectralEnvelope(nr = 512, nc = 50, formants = NA,
  temperature = .1, vocalTract = 15.5)))

# no formants at all
image(t(getSpectralEnvelope(nr = 512, nc = 50,
  formants = NA, temperature = 0)))

# manual specification of formants
image(t(getSpectralEnvelope(nr = 512, nc = 50,
samplingRate = 16000, formants = list(
  'f1' = data.frame('time' = 0, 'freq' = 900, 'amp' = 30, 'width' = 120),
  'f2' = data.frame('time' = 0, 'freq' = 1300, 'amp' = 30, 'width' = 120),
  'f3' = data.frame('time' = 0, 'freq' = 3200, 'amp' = 20, 'width' = 200))))))

```

 HzToSemitones

Convert Hz to semitones

Description

Converts from Hz to semitones above C0 (~16.4 Hz). This may not seem very useful, but note that (1) this gives you a nice logarithmic scale for generating natural pitch transitions, (2) with the added benefit of getting musical notation for free from notesDict (see examples).

Usage

```
HzToSemitones(h)
```

Arguments

h vector or matrix of frequencies (Hz)

Examples

```

s = HzToSemitones(c(440, 293, 115))
# to convert to musical notation
notesDict$note[1 + round(s)]
# note the "1 +": semitones ABOVE C0, i.e. notesDict[1, ] is C0

```

matchPars	<i>Match soundgen pars (experimental)</i>
-----------	---

Description

Attempts to find settings for [soundgen](#) that will reproduce an existing sound. The principle is to mutate control parameters, trying to improve fit to target. The currently implemented optimization algorithm is simple hill climbing. Disclaimer: this function is experimental and may or may not work for particular tasks. It is intended as a supplement to - not replacement of - manual optimization. See the vignette on sound generation for more information.

Usage

```
matchPars(target, samplingRate = NULL, pars = NULL, init = NULL,
  method = c("cor", "cosine", "pixel", "dtw"), probMutation = 0.25,
  stepVariance = 0.1, maxIter = 50, minExpectedDelta = 0.001,
  windowLength = 40, overlap = 50, step = NULL, verbose = TRUE,
  padWith = NA, penalizeLengthDif = TRUE, throwaway = -120,
  maxFreq = NULL)
```

Arguments

target	the sound we want to reproduce using soundgen: path to a .wav file or numeric vector
samplingRate	sampling rate of target (only needed if target is a numeric vector, rather than a .wav file)
pars	arguments to soundgen that we are attempting to optimize
init	a list of initial values for the optimized parameters pars and the values of other arguments to soundgen that are fixed at non-default values (if any)
method	method of comparing mel-transformed spectra of two sounds: "cor" = average Pearson's correlation of mel-transformed spectra of individual FFT frames; "cosine" = same as "cor" but with cosine similarity instead of Pearson's correlation; "pixel" = absolute difference between each point in the two spectra; "dtw" = discrete time warp with dtw
probMutation	the probability of a parameter mutating per iteration
stepVariance	scale factor for calculating the size of mutations
maxIter	maximum number of mutated sounds produced without improving the fit to target
minExpectedDelta	minimum improvement in fit to target required to accept the new sound candidate
windowLength	length of FFT window, ms
overlap	overlap between successive FFT frames, %
step	you can override overlap by specifying FFT step, ms

verbose	if TRUE, plays back the accepted candidate at each iteration and reports the outcome
padWith	compared spectra are padded with either silence (<code>padWith = 0</code>) or with NA's (<code>padWith = NA</code>) to have the same number of columns. When the sounds are of different duration, padding with zeros rather than NA's improves the fit to target measured by <code>method = 'pixel'</code> and <code>'dtw'</code> , but it has no effect on <code>'cor'</code> and <code>'cosine'</code> .
penalizeLengthDif	if TRUE, sounds of different length are considered to be less similar; if FALSE, only the overlapping parts of two sounds are compared
throwaway	parts of the spectra quieter than <code>throwaway</code> dB are not compared
maxFreq	parts of the spectra above <code>maxFreq</code> Hz are not compared

Value

Returns a list of length 2: `$history` contains the tried parameter values together with their fit to target (`$history$sim`), and `$pars` contains a list of the final - hopefully the best - parameter settings.

Examples

```
## Not run:
playback = c(TRUE, FALSE)[2] # set to TRUE to play back the audio from examples

target = soundgen(repeatBout = 3, syllLen = 120, pauseLen = 70,
  pitchAnchors = data.frame(time = c(0, 1), value = c(300, 200)),
  rolloff = -5, play = playback) # we hope to reproduce this sound

# Match pars based on acoustic analysis alone, without any optimization.
# This *MAY* match temporal structure, pitch, and stationary formants
m1 = matchPars(target = target,
  samplingRate = 16000,
  maxIter = 0, # no optimization, only acoustic analysis
  verbose = playback)
cand1 = do.call(soundgen, c(m1$pars, list(play = playback, temperature = 0)))

# Try to improve the match by optimizing rolloff
# (this may take a few minutes to run, and the results may vary)
m2 = matchPars(target = target,
  samplingRate = 16000,
  pars = 'rolloff',
  maxIter = 100,
  verbose = playback)
# rolloff should be moving from default (-12) to target (-5):
sapply(m2$history, function(x) x$pars$rolloff)
cand2 = do.call(soundgen, c(m2$pars, list(play = playback, temperature = 0)))

## End(Not run)
```

 morph

Morph sounds

Description

Takes two formulas for synthesizing two target sounds with [soundgen](#) and produces a number of intermediate forms (morphs), attempting to go from one target sound to the other in a specified number of equal steps.

Usage

```
morph(formula1, formula2, nMorphs, playMorphs = TRUE, savePath = NA,
      samplingRate = 16000)
```

Arguments

formula1, formula2	lists of parameters for calling soundgen that produce the two target sounds between which morphing will occur. Character strings containing the full call to soundgen are also accepted (see examples)
nMorphs	the length of morphing sequence, including target sounds
playMorphs	if TRUE, the morphing sequence will be played
savePath	if it is the path to an existing directory, morphs will be saved there as individual .wav files (defaults to NA)
samplingRate	sampling rate of output, Hz. NB: must be the same as in formula1 and formula2!

Value

A list of two sublists (`$formulas` and `$sounds`), each sublist of length `nMorphs`. For ex., the formula for the second hybrid is `m$formulas[[2]]`, and the waveform is `m$sounds[[2]]`

Examples

```
# write two formulas or copy-paste them from soundgen_app() or presets, for example:
m = morph(formula1 = list(repeatBout = 2),
          # equivalently: formula1 = 'soundgen(repeatBout = 2)',
          formula2 = presets$Misc$Dog_bark,
          nMorphs = 5, playMorphs = FALSE)
# use $formulas to access formulas for each morph, $sounds for waveforms
# m$formulas[[4]]
# playme(m$sounds[[3]])
```

notesDict	<i>Conversion table from Hz to semitones above C0 to musical notation</i>
-----------	---

Description

A dataframe of 132 rows and 2 columns: "note" and "freq" (Hz)

Usage

```
notesDict
```

Format

An object of class `data.frame` with 132 rows and 2 columns.

Examples

```
# To recompile:
notes = c('C', 'C\u266F', 'D', 'D\u266F', 'E', 'F', 'F\u266F', 'G', 'G\u266F', 'A', 'B\u266D', 'B')
nOct = 11
notes_all = paste0(notes, rep(0:(nOct - 1), each = 12))
# 440 / 32 = 13.75 # A-1, and C0 is 3 semitones higher: 16.3516 Hz exactly.
c0 = 13.75 * 2 ^ (3 / 12)
notes_freq = round (c0 * 2^(0:(12 * nOct - 1) / 12), 1)
```

optimizePars	<i>Optimize parameters for acoustic analysis</i>
--------------	--

Description

This customized wrapper for `optim` attempts to optimize the parameters of `segmentFolder` or `analyzeFolder` by comparing the results with a manually annotated "key". This optimization function uses a single measurement per audio file (e.g., median pitch or the number of syllables). For other purposes, you may want to adapt the optimization function so that the key specifies the exact timing of syllables, their median length, frame-by-frame pitch values, or any other characteristic that you want to optimize for. The general idea remains the same, however: we want to tune function parameters to fit our type of audio and research priorities. The default settings of `segmentFolder` and `analyzeFolder` have been optimized for human non-linguistic vocalizations.

Usage

```
optimizePars(myfolder, key, myfun, pars, bounds = NULL, fitnessPar,
  fitnessFun = function(x) 1 - cor(x, key, use = "pairwise.complete.obs"),
  nIter = 10, init = NULL, initSD = 0.2, control = list(maxit = 50,
  reltol = 0.01, trace = 0), otherPars = list(plot = FALSE, verbose = FALSE),
  mygrid = NULL, verbose = TRUE)
```


Arguments

myfolder	path to where the .wav files live
key	a vector containing the "correct" measurement that we are aiming to reproduce
myfun	the function being optimized: either 'segmentFolder' or 'analyzeFolder' (in quotes)
pars	names of arguments to myfun that should be optimized
bounds	a list setting the lower and upper boundaries for possible values of optimized parameters. For ex., if we optimize smooth and smoothOverlap, reasonable bounds might be list(low = c(5, 0), high = c(500, 95))
fitnessPar	the name of output variable that we are comparing with the key, e.g. 'nBursts' or 'pitch_median'
fitnessFun	the function used to evaluate how well the output of myfun fits the key. Defaults to 1 - Pearson's correlation (i.e. 0 is perfect fit, 1 is awful fit). For pitch, log scale is more meaningful, so a good fitness criterion is <code>function(x) 1 - cor(log(x), log(key), use = 'pairwise.complete.obs')</code>
nIter	repeat the optimization several times to check convergence
init	initial values of optimized parameters (if NULL, the default values are taken from the definition of myfun)
initSD	each optimization begins with a random seed, and initSD specifies the SD of normal distribution used to generate random deviation of initial values from the defaults
control	a list of control parameters passed on to <code>optim</code> . The method used is "Nelder-Mead"
otherPars	a list of additional arguments to myfun
mygrid	a dataframe with one column per parameter to optimize, with each row specifying the values to try. If not NULL, optimizePars simply evaluates each combination of parameter values, without calling <code>optim</code> (see examples)
verbose	if TRUE, reports the values of parameters evaluated and fitness

Details

If your sounds are very different from human non-linguistic vocalizations, you may want to change the default values of other arguments to speed up convergence. Adapt the code to enforce suitable constraints, depending on your data.

Value

Returns a matrix with one row per iteration with fitness in the first column and the best values of each of the optimized parameters in the remaining columns.

Examples

```
## Not run:
# download 260 sounds from Anikin & Persson (2017)
# http://cogsci.se/personal/results/
```

```

# 01_anikin-persson_2016_naturalistics-non-linguistic-vocalizations/260sounds_wav.zip
# unzip them into a folder, say '~/Downloads/temp'
myfolder = '~/Downloads/temp' # 260 .wav files live here

# Optimization of SEGMENTATION
# import manual counts of syllables in 260 sounds from Anikin & Persson (2017) (our "key")
key = segmentManual # a vector of 260 integers
# run optimization loop several times with random initial values to check convergence
# NB: with 260 sounds and default settings, this might take ~20 min per iteration!
res = optimizePars(myfolder = myfolder, myfun = 'segmentFolder', key = key,
  pars = c('shortestSyl', 'shortestPause', 'sylThres'),
  fitnessPar = 'nBursts',
  nIter = 3, control = list(maxit = 50, reltol = .01, trace = 0))

# examine the results
print(res)
for (c in 2:ncol(res)) {
  plot(res[, c], res[, 1], main = colnames(res)[c])
}
pars = as.list(res[1, 2:ncol(res)]) # top candidate (best pars)
s = do.call(segmentFolder, c(myfolder, pars)) # segment with best pars
cor(key, as.numeric(s[, fitnessPar]))
boxplot(as.numeric(s[, fitnessPar]) ~ as.integer(key), xlab='key')
abline(a=0, b=1, col='red')

# Try a grid with particular parameter values instead of formal optimization
res = optimizePars(myfolder = myfolder, myfun = 'segmentFolder', key = segment_manual,
  pars = c('shortestSyl', 'shortestPause'),
  fitnessPar = 'nBursts',
  mygrid = expand.grid(shortestSyl = c(30, 40),
    shortestPause = c(30, 40, 50)))
1 - res$fit # correlations with key

# Optimization of PITCH TRACKING (takes several hours!)
res = optimizePars(myfolder = myfolder,
  myfun = 'analyzeFolder',
  key = log(pitchManual), # log-scale better for pitch
  pars = c('specThres', 'specSmooth'),
  bounds = list(low = c(0, 0), high = c(1, Inf)),
  fitnessPar = 'pitch_median',
  nIter = 2,
  otherPars = list(plot = FALSE, verbose = FALSE, step = 50,
    pitchMethods = 'spec'),
  fitnessFun = function(x) {
    1 - cor(log(x), key, use = 'pairwise.complete.obs') *
    (1 - mean(is.na(x) & !is.na(key))) # penalize failing to detect F0
  })

## End(Not run) # end of dontrun

```

Description

A dataset containing defaults and ranges of key variables in the Shiny app. Adjust as needed.

Usage

```
permittedValues
```

Format

A matrix with 58 rows and 4 variables:

default default value

low lowest permitted value

high highest permitted value

step increment for adjustment ...

pitchManual

Manual pitch estimation in 260 sounds

Description

A vector of manually verified pitch values per sound in the corpus of 590 human non-linguistic emotional vocalizations from Anikin & Persson (2017). The corpus can be downloaded from http://cogsci.se/personal/results/01_anikin-persson_2016_naturalistics-non-linguistic-vocalizations/01_anikin-persson_2016_naturalistic-non-linguistic-vocalizations.html

Usage

```
pitchManual
```

Format

An object of class `numeric` of length 260.

playme	<i>Play audio</i>
--------	-------------------

Description

Plays an audio file or a numeric vector. This is a simple wrapper for the functionality provided by [play](#)

Usage

```
playme(sound, samplingRate = 16000)
```

Arguments

sound	a vector of numbers on any scale or a path to a .wav file
samplingRate	sampling rate (only needed if sound is a vector)

Examples

```
## Not run: playme('~/myfile.wav')
f0_Hz = 440
sound = sin(2 * pi * f0_Hz * (1:16000) / 16000)
# playme(sound, 16000)
# in case of errors, look into tuneR::play()
```

presets	<i>Presets</i>
---------	----------------

Description

A library of presets for easy generation of a few nice sounds.

Usage

```
presets
```

Format

A list of length 4.

segment	<i>Segment a sound</i>
---------	------------------------

Description

Finds syllables and bursts. Syllables are defined as continuous segments with amplitude above threshold. Bursts are defined as local maxima in amplitude envelope that are high enough both in absolute terms (relative to the global maximum) and with respect to the surrounding region (relative to local minima). See the vignette on acoustic analysis for details.

Usage

```
segment(x, samplingRate = NULL, windowLength = 40, overlap = 80,
        shortestSyl = 40, shortestPause = 40, sylThres = 0.9,
        interburst = NULL, interburstMult = 1, burstThres = 0.075,
        peakToTrough = 3, troughLeft = TRUE, troughRight = FALSE,
        summary = FALSE, plot = FALSE, savePath = NA, ...)
```

Arguments

<code>x</code>	path to a .wav file or a vector of amplitudes with specified <code>samplingRate</code>
<code>samplingRate</code>	sampling rate of <code>x</code> (only needed if <code>x</code> is a numeric vector, rather than a .wav file)
<code>windowLength</code> , <code>overlap</code>	length (ms) and overlap (window used to produce the amplitude envelope, see env)
<code>shortestSyl</code>	minimum acceptable length of syllables, ms
<code>shortestPause</code>	minimum acceptable break between syllables, ms. Syllables separated by less time are merged. To avoid merging, specify <code>shortestPause = NA</code>
<code>sylThres</code>	amplitude threshold for syllable detection (as a proportion of global mean amplitude of smoothed envelope)
<code>interburst</code>	minimum time between two consecutive bursts (ms). If specified, it overrides <code>interburstMult</code>
<code>interburstMult</code>	multiplier of the default minimum interburst interval (median syllable length or, if no syllables are detected, the same number as <code>shortestSyl</code>). Only used if <code>interburst</code> is not specified. Larger values improve detection of unusually broad shallow peaks, while smaller values improve the detection of sharp narrow peaks
<code>burstThres</code>	to qualify as a burst, a local maximum has to be at least <code>burstThres</code> times the height of the global maximum of amplitude envelope
<code>peakToTrough</code>	to qualify as a burst, a local maximum has to be at least <code>peakToTrough</code> times the local minimum on the LEFT over analysis window (which is controlled by <code>interburst</code> or <code>interburstMult</code>)
<code>troughLeft</code> , <code>troughRight</code>	should local maxima be compared to the trough on the left and/or right of it? Default to TRUE and FALSE, respectively

summary	if TRUE, returns only a summary of the number and spacing of syllables and vocal bursts. If FALSE, returns a list containing full stats on each syllable and bursts (location, duration, amplitude, ...)
plot	if TRUE, produces a segmentation plot
savePath	full path to the folder in which to save the plots. Defaults to NA
...	other graphical parameters passed to <code>plot</code>

Details

The algorithm is very flexible, but the parameters may be hard to optimize by hand. If you have an annotated sample of the sort of audio you are planning to analyze, with syllables and/or bursts counted manually, you can use it for automatic optimization of control parameters (see [optimizePars](#)). The defaults are the results of just such optimization against 260 human vocalizations in Anikin, A. & Persson, T. (2017). Non-linguistic vocalizations from online amateur videos for emotion research: a validated corpus. *Behavior Research Methods*, 49(2): 758-771.

Value

If `summary = TRUE`, returns only a summary of the number and spacing of syllables and vocal bursts. If `summary = FALSE`, returns a list containing full stats on each syllable and bursts (location, duration, amplitude, ...).

Examples

```
sound = soundgen(nSyl = 8, syllLen = 50, pauseLen = 70,
  pitchAnchors = list(time = c(0, 1), value = c(368, 284)), temperature = 0.1,
  noiseAnchors = list(time = c(0, 67, 86, 186), value = c(-45, -47, -89, -120)),
  rolloff_noise = -8, amplAnchorsGlobal = list(time = c(0, 1), value = c(120, 20)))
spectrogram(sound, samplingRate = 16000, osc = TRUE)
# playme(sound, samplingRate = 16000)

s = segment(sound, samplingRate = 16000, plot = TRUE)
# accept quicker and quieter syllables
s = segment(sound, samplingRate = 16000, plot = TRUE,
  shortestSyl = 25, shortestPause = 25, sylThres = .6)
# look for narrower, sharper bursts
s = segment(sound, samplingRate = 16000, plot = TRUE,
  shortestSyl = 25, shortestPause = 25, sylThres = .6,
  interburstMult = 1)

# just a summary
segment(sound, samplingRate = 16000, summary = TRUE)
```

segmentFolder

Segment all files in a folder

Description

Finds syllables and bursts in all .wav files in a folder.

Usage

```
segmentFolder(myfolder, shortestSyl = 40, shortestPause = 40,
  sylThres = 0.9, interburst = NULL, interburstMult = 1,
  burstThres = 0.075, peakToTrough = 3, troughLeft = TRUE,
  troughRight = FALSE, windowLength = 40, overlap = 80, summary = TRUE,
  plot = FALSE, savePath = NA, verbose = TRUE, reportEvery = 10, ...)
```

Arguments

myfolder	full path to target folder
shortestSyl	minimum acceptable length of syllables, ms
shortestPause	minimum acceptable break between syllables, ms. Syllables separated by less time are merged. To avoid merging, specify shortestPause = NA
sylThres	amplitude threshold for syllable detection (as a proportion of global mean amplitude of smoothed envelope)
interburst	minimum time between two consecutive bursts (ms). If specified, it overrides interburstMult
interburstMult	multiplier of the default minimum interburst interval (median syllable length or, if no syllables are detected, the same number as shortestSyl). Only used if interburst is not specified. Larger values improve detection of unusually broad shallow peaks, while smaller values improve the detection of sharp narrow peaks
burstThres	to qualify as a burst, a local maximum has to be at least burstThres times the height of the global maximum of amplitude envelope
peakToTrough	to qualify as a burst, a local maximum has to be at least peakToTrough times the local minimum on the LEFT over analysis window (which is controlled by interburst or interburstMult)
troughLeft	should local maxima be compared to the trough on the left and/or right of it? Default to TRUE and FALSE, respectively
troughRight	should local maxima be compared to the trough on the left and/or right of it? Default to TRUE and FALSE, respectively
windowLength	length (ms) and overlap (window used to produce the amplitude envelope, see env)
overlap	length (ms) and overlap (window used to produce the amplitude envelope, see env)
summary	if TRUE, returns only a summary of the number and spacing of syllables and vocal bursts. If FALSE, returns a list containing full stats on each syllable and bursts (location, duration, amplitude, ...)
plot	if TRUE, produces a segmentation plot
savePath	full path to the folder in which to save the plots. Defaults to NA
verbose, reportEvery	if TRUE, reports progress every reportEvery files and estimated time left
...	other graphical parameters passed to plot

Details

This is just a convenient wrapper for `segment` intended for analyzing the syllables and bursts in a large number of audio files at a time. In verbose mode, it also reports ETA every ten iterations. With default settings, running time should be about a second per minute of audio.

Value

If `summary` is `TRUE`, returns a dataframe with one row per audio file. If `summary` is `FALSE`, returns a list of detailed descriptives.

Examples

```
## Not run:
# download 260 sounds from Anikin & Persson (2017)
# http://cogsci.se/personal/results/
# 01_anikin-persson_2016_naturalistics-non-linguistic-vocalizations/260sounds_wav.zip
# unzip them into a folder, say '~/Downloads/temp'
myfolder = '~/Downloads/temp' # 260 .wav files live here
s = segmentFolder(myfolder, verbose = TRUE)

# Check accuracy: import a manual count of syllables (our "key")
key = segmentManual # a vector of 260 integers
trial = as.numeric(s$nBursts)
cor(key, trial, use = 'pairwise.complete.obs')
boxplot(trial ~ as.integer(key), xlab='key')
abline(a=0, b=1, col='red')

## End(Not run)
```

segmentManual

Manual counts of syllables in 260 sounds

Description

A vector of the number of syllables in the corpus of 260 human non-linguistic emotional vocalizations from Anikin & Persson (2017). The corpus can be downloaded from http://cogsci.se/personal/results/01_anikin-persson_2016_naturalistics-non-linguistic-vocalizations/01_anikin-persson_2016_naturalistic-non-linguistic-vocalizations.html

Usage

```
segmentManual
```

Format

An object of class `numeric` of length 260.

semitonesToHz	<i>Convert semitones to Hz</i>
---------------	--------------------------------

Description

Converts from semitones above C0 (~16.4 Hz) to Hz. See [HzToSemi tones](#)

Usage

```
semitonesToHz(s)
```

Arguments

s	vector or matrix of frequencies (semitones above C0)
---	--

soundgen	<i>Generate a sound</i>
----------	-------------------------

Description

Generates a bout of one or more syllables with pauses between them. Two basic components are synthesized: the harmonic component (the sum of sine waves with frequencies that are multiples of the fundamental frequency) and the noise component. Both components can be filtered with independently specified formants. Intonation and amplitude contours can be applied both within each syllable and across multiple syllables. Suggested application: synthesis of animal or human non-linguistic vocalizations. For more information, see <http://cogsci.se/soundgen.html> and the vignette on sound generation.

Usage

```
soundgen(repeatBout = 1, nSyl = 1, syllLen = 300, pauseLen = 200,
  pitchAnchors = data.frame(time = c(0, 0.1, 0.9, 1), value = c(100, 150, 135,
  100)), pitchAnchorsGlobal = NA, temperature = 0.025,
  tempEffects = list(syllLenDep = 0.02, formDrift = 0.3, formDisp = 0.2,
  pitchDriftDep = 0.5, pitchDriftFreq = 0.125, pitchAnchorsDep = 0.05,
  noiseAnchorsDep = 0.1, amplAnchorsDep = 0.1), maleFemale = 0,
  creakyBreathy = 0, nonlinBalance = 0, nonlinDep = 50, jitterLen = 1,
  jitterDep = 3, vibratoFreq = 5, vibratoDep = 0, shimmerDep = 0,
  attackLen = 50, rolloff = -12, rolloffOct = -12, rolloffKHz = -6,
  rolloffParab = 0, rolloffParabHarm = 3, rolloffLip = 6,
  formants = list(f1 = list(time = 0, freq = 860, amp = 30, width = 120), f2 =
  list(time = 0, freq = 1280, amp = 40, width = 120), f3 = list(time = 0, freq =
  2900, amp = 25, width = 200)), formantDep = 1, formantDepStoch = 30,
  vocalTract = 15.5, subFreq = 100, subDep = 100, shortestEpoch = 300,
  amDep = 0, amFreq = 30, amShape = 0, noiseAnchors = data.frame(time =
```

```
c(0, 300), value = c(-120, -120)), formantsNoise = NA, rolloffNoise = -14,
mouthAnchors = data.frame(time = c(0, 1), value = c(0.5, 0.5)),
amplAnchors = NA, amplAnchorsGlobal = NA, samplingRate = 16000,
windowLength = 50, overlap = 75, addSilence = 100, pitchFloor = 50,
pitchCeiling = 3500, pitchSamplingRate = 3500, throwaway = -120,
invalidArgAction = c("adjust", "abort", "ignore")[1], plot = FALSE,
play = FALSE, savePath = NA, ...)
```

Arguments

repeatBout	the number of times the whole bout should be repeated
nSyl	the number of syllables in the bout. Intonation, amplitude, and formants contours span multiple syllables, but not multiple bouts (see Details)
syllLen	average duration of each syllable, ms
pauseLen	average duration of pauses between syllables, ms
pitchAnchors	a numeric vector of f0 values in Hz (assuming equal time steps) or a dataframe specifying the time (ms) and value (Hz) of each anchor. These anchors are used to create a smooth contour of fundamental frequency f0 (pitch) within one syllable (see Examples)
pitchAnchorsGlobal	unlike pitchAnchors, these anchors are used to create a smooth contour of average f0 across multiple syllables. The values are in semitones relative to the existing pitch, i.e. 0 = no change
temperature	hyperparameter for regulating the amount of stochasticity in sound generation
tempEffects	a list of scale factors regulating the effect of temperature on particular parameters. To change, specify just those pars that you want to modify, don't rewrite the whole list (defaults are hard-coded). syllLenDep: random variation of the duration of syllables and pauses; formDrift: the amount of random drift of formants; formDisp: irregularity of the dispersion of stochastic formants; pitchDriftDep: amount of slow random drift of f0; pitchDriftFreq: frequency of slow random drift of f0; pitchAnchorsDep, noiseAnchorsDep, amplAnchorsDep: random fluctuations of user-specified pitch / noise / amplitude anchors
maleFemale	hyperparameter for shifting f0 contour, formants, and vocalTract to make the speaker appear more male (-1...0) or more female (0...+1)
creakyBreathy	hyperparameter for a rough adjustment of voice quality from creaky (-1) to breathy (+1)
nonlinBalance	hyperparameter for regulating the (approximate) proportion of sound with different regimes of pitch effects (none / subharmonics only / subharmonics and jitter). 0% = no noise; 100% = the entire sound has jitter + subharmonics. Ignored if temperature = 0
nonlinDep	hyperparameter for regulating the intensity of subharmonics and jitter, 0 to 100% (50% = jitter and subharmonics are as specified, <50% weaker, >50% stronger). Ignored if temperature = 0
jitterLen	duration of stable periods between pitch jumps, ms. Use a low value for harsh noise, a high value for irregular vibrato or shaky voice

jitterDep	cycle-to-cycle random pitch variation, semitones
vibratoFreq	the rate of regular pitch modulation, or vibrato, Hz
vibratoDep	the depth of vibrato, semitones
shimmerDep	random variation in amplitude between individual glottal cycles (0 to 100% of original amplitude of each cycle)
attackLen	duration of fade-in / fade-out at each end of syllables and noise (ms)
rolloff	basic rolloff at a constant rate of rolloff dB/octave (exponential decay). See getRolloff for more details
rolloffOct	basic rolloff changes from lower to upper harmonics (regardless of f0) by rolloffOct dB/oct. For example, we can get steeper rolloff in the upper part of the spectrum
rolloffKHz	rolloff changes linearly with f0 by rolloffKHz dB/kHz. For ex., -6 dB/kHz gives a 6 dB steeper basic rolloff as f0 goes up by 1000 Hz
rolloffParab	an optional quadratic term affecting only the first rolloffParabHarm harmonics. The middle harmonic of the first rolloffParabHarm harmonics is amplified or dampened by rolloffParab dB relative to the basic exponential decay.
rolloffParabHarm	the number of harmonics affected by rolloffParab
rolloffLip	the effect of lip radiation on source spectrum, dB/oct (the default of +6 dB/oct produces a high-frequency boost when the mouth is open)
formants	either a character string like "aui" referring to default presets for speaker "M1" or a list of formant times, frequencies, amplitudes, and bandwidths (see ex. below). formants = NA defaults to schwa. Time stamps for formants and mouthOpening can be specified in ms or an any other arbitrary scale. See getSpectralEnvelope for more details
formantDep	scale factor of formant amplitude (1 = no change relative to amplitudes in formants)
formantDepStoch	the amplitude of additional stochastic formants added above the highest specified formant, dB (only if temperature > 0)
vocalTract	the length of vocal tract, cm. Used for calculating formant dispersion (for adding extra formants) and formant transitions as the mouth opens and closes
subFreq	target frequency of subharmonics, Hz (lower than f0, adjusted dynamically so f0 is always a multiple of subFreq)
subDep	the width of subharmonic band, Hz. Regulates how quickly the strength of subharmonics fades as they move away from harmonics in f0 stack. Low values produce narrow sidebands, high values produce uniformly strong subharmonics
shortestEpoch	minimum duration of each epoch with unchanging subharmonics regime, in ms
amDep	amplitude modulation depth, modulation with amplitude range equal to the dynamic range of the sound
amFreq	amplitude modulation frequency, Hz
amShape	amplitude modulation shape (-1 to +1, defaults to 0)
noiseAnchors	a numeric vector of noise amplitudes (-120 dB = none, 0 dB = as loud as voiced component) or a dataframe specifying the time (ms) and amplitude (dB) of anchors for generating the noise component such as aspiration, hissing, etc

formantsNoise	the same as formants, but for the noise component instead of the harmonic component. If NA (default), the noise component will be filtered through the same formants as the harmonic component, approximating aspiration noise [h]
rolloffNoise	rolloff of noise, dB/octave. It is analogous to rolloff, but while rolloff applies to the harmonic component, rolloffNoise applies to the noise component
mouthAnchors	a numeric vector of mouth opening (0 to 1, 0.5 = neutral, i.e. no modification) or a dataframe specifying the time (ms) and value of mouth opening
amplAnchors	a numeric vector of amplitude envelope (0 to 1) or a dataframe specifying the time (ms) and value of amplitude anchors
amplAnchorsGlobal	a numeric vector of global amplitude envelope spanning multiple syllables or a dataframe specifying the time (ms) and value (0 to 1) of each anchor
samplingRate	sampling frequency, Hz
windowLength	length of FFT window, ms
overlap	FFT window overlap, %
addSilence	silence before and after the bout, ms
pitchFloor, pitchCeiling	lower & upper bounds of f0
pitchSamplingRate	sampling frequency of the pitch contour only, Hz. Low values reduce processing time. A rule of thumb is to set this to the same value as pitchCeiling
throwaway	discard harmonics and noise that are quieter than this number (in dB, defaults to -120) to save computational resources
invalidArgAction	what to do if an argument is invalid or outside the range in permittedValues: 'adjust' = reset to default value, 'abort' = stop execution, 'ignore' = throw a warning and continue (may crash)
plot	if TRUE, plots a spectrogram
play	if TRUE, plays the synthesized sound. In case of errors, try setting another default player for play
savePath	full path for saving the output, e.g. '~/Downloads/temp.wav'. If NA (default), doesn't save anything
...	other plotting parameters passed to spectrogram

Value

Returns the synthesized waveform as a numeric vector.

Examples

```
## Not run:
# NB: GUI for soundgen is available as a Shiny app.
# Type "soundgen_app()" to start it

playback = c(TRUE, FALSE)[2] # set to TRUE to play back the audio from examples
```

```

sound = soundgen(play = playback)
# spectrogram(sound, 16000, osc = TRUE)
# playme(sound)

# Use the in-built collection of presets:
# names(presets) # speakers
# names(presets$Chimpanzee) # calls per speaker
s1 = eval(parse(text = presets$Chimpanzee$Scream_conflict)) # screaming chimp
# playme(s1)
s2 = eval(parse(text = presets$F1$Scream_conflict))
# playme(s2)
# unless temperature is 0, the sound is different every time
for (i in 1:3) sound = soundgen(play = playback, temperature = .2)

# Bouts versus syllables. Compare:
sound = soundgen(formants = 'uai', repeatBout = 3, play = playback)
sound = soundgen(formants = 'uai', nSyl = 3, play = playback)

# Intonation contours per syllable and globally:
sound = soundgen(nSyl = 5, sylLen = 200, pauseLen = 140,
  play = playback, pitchAnchors = data.frame(
    time = c(0, 0.65, 1), value = c(977, 1540, 826)),
  pitchAnchorsGlobal = data.frame(time = c(0, .5, 1), value = c(-6, 7, 0)))

# Subharmonics in sidebands (noisy scream)
sound = soundgen (nonlinBalance = 100, subFreq = 75, subDep = 130,
  pitchAnchors = data.frame(
    time = c(0, .3, .9, 1), value = c(1200, 1547, 1487, 1154)),
  sylLen = 800,
  play = playback, plot = TRUE)

# Jitter and mouth opening (bark, dog-like)
sound = soundgen(repeatBout = 2, sylLen = 160, pauseLen = 100,
  nonlinBalance = 100, subFreq = 100, subDep = 60, jitterDep = 1,
  pitchAnchors = data.frame(time = c(0, 0.52, 1), value = c(559, 785, 557)),
  mouthAnchors = data.frame(time = c(0, 0.5, 1), value = c(0, 0.5, 0)),
  vocalTract = 5, play = playback)

## End(Not run)

```

soundgen_app

Soundgen shiny app

Description

Starts a shiny app, which provides an interactive wrapper to [soundgen](#)

Usage

```
soundgen_app()
```

spectrogram	<i>Spectrogram</i>
-------------	--------------------

Description

Produces the spectrogram of a sound using short-term Fourier transform. This is a simplified version of `spectro` with fewer plotting options, but with added routines for noise reduction, smoothing in time and frequency domains, and controlling contrast and brightness.

Usage

```
spectrogram(x, samplingRate = NULL, windowLength = 50, step = NULL,
  overlap = 70, wn = "gaussian", zp = 0, smoothFreq = 0,
  smoothTime = 0, qTime = 0, percentNoise = 10, noiseReduction = 0,
  contrast = 0.2, brightness = 0, method = c("spectrum",
  "spectralDerivative")[1], output = c("none", "original", "processed")[1],
  ylim = NULL, plot = TRUE, osc = FALSE, colorTheme = c("bw", "seewave",
  "...")[1], xlab = "", frameBank = NULL, duration = NULL, ...)
```

Arguments

<code>x</code>	path to a .wav file or a vector of amplitudes with specified <code>samplingRate</code>
<code>samplingRate</code>	sampling rate of <code>x</code> (only needed if <code>x</code> is a numeric vector, rather than a .wav file)
<code>windowLength</code>	length of FFT window, ms
<code>step</code>	you can override <code>overlap</code> by specifying FFT step, ms
<code>overlap</code>	overlap between successive FFT frames, %
<code>wn</code>	window type: gaussian, hanning, hamming, bartlett, rectangular, blackman, flat-top
<code>zp</code>	window length after zero padding, points
<code>smoothFreq</code> , <code>smoothTime</code>	length of the window, in data points (0 to +inf), for calculating a rolling median. Applies median smoothing to spectrogram in frequency and time domains, respectively
<code>qTime</code>	the quantile to be subtracted for each frequency bin. For ex., if <code>qTime = 0.5</code> , the median of each frequency bin (over the entire sound duration) will be calculated and subtracted from each frame (see examples)
<code>percentNoise</code>	percentage of frames (0 to 100%) used for calculating noise spectrum
<code>noiseReduction</code>	how much noise to remove (0 to +inf, recommended 0 to 2). 0 = no noise reduction, 2 = strong noise reduction: $spectrum - (noiseReduction * noiseSpectrum)$, where <code>noiseSpectrum</code> is the average spectrum of frames with entropy exceeding the quantile set by <code>percentNoise</code>
<code>contrast</code>	spectrum is exponentiated by contrast (-inf to +inf, recommended -1 to +1). Contrast >0 increases sharpness, <0 decreases sharpness

brightness	how much to "lighten" the image (>0 = lighter, <0 = darker)
method	plot spectrum ('spectrum') or the first time derivative of the spectrum ('spectralDerivative')
output	specifies what to return: nothing ('none'), unmodified spectrogram ('original'), or denoised and/or smoothed spectrogram ('processed')
ylim	frequency range to plot, kHz (defaults to 0 to Nyquist frequency)
plot	should a spectrogram be plotted? TRUE / FALSE
osc	should an oscillogram be shown under the spectrogram? TRUE / FALSE
colorTheme	black and white ('bw'), as in seewave package ('seewave'), or another color theme (e.g. 'heat.colors')
xlab	label for x-axis
frameBank	ignore (only needed for pitch tracking)
duration	ignore (only needed for pitch tracking)
...	other graphical parameters passed to <code>seewave::filled.contour.modif2</code>

Value

Returns nothing (if output = 'none'), raw spectrum (if output = 'original'), denoised and/or smoothed spectrum (if output = 'processed'), or spectral derivatives (if method = 'spectralDerivative') as a matrix of real numbers.

Examples

```
## Not run:
# synthesize a sound 1 s long, with gradually increasing hissing noise
sound = soundgen(syllen = 1000, temperature = 0, noiseAnchors = list(
  time = c(0, 1300), value = c(-120, 0)), formantsNoise = list(
  f1 = list(time = 0, freq = 5000, width = 10000, amp = 0)))
# playme(sound, samplingRate = 16000)

# basic spectrogram
spectrogram(sound, samplingRate = 16000)
# add an oscillogram
spectrogram(sound, samplingRate = 16000, osc = TRUE)
# broad-band instead of narrow-band
spectrogram(sound, samplingRate = 16000, windowLength = 5)
# spectral derivatives
spectrogram(sound, samplingRate = 16000, method = 'spectralDerivative')

# focus only on values in the upper 5% for each frequency bin
spectrogram(sound, samplingRate = 16000, qTime = 0.95)

# detect 10% of the noisiest frames based on entropy and remove the pattern
# found in those frames (in this cases, breathing)
spectrogram(sound, samplingRate = 16000, noiseReduction = 1.1,
  brightness = -2) # white noise almost gone

# apply median smoothing in both time and frequency domains
```

```
spectrogram(sound, samplingRate = 16000, smoothFreq = 5,
  smoothTime = 5)

# increase contrast, reduce brightness
spectrogram(sound, samplingRate = 16000, contrast = 1, brightness = -1)

# add bells and whistles
spectrogram(sound, samplingRate = 16000, osc = TRUE, noiseReduction = 1.1,
  brightness = -1, colorTheme = 'heat.colors', xlab = 'Time, ms',
  ylab = 'Frequency, kHz', ylim = c(0,5))

## End(Not run)
```

 ssm

Self-similarity matrix

Description

Calculates the self-similarity matrix and novelty vector of a sound.

Usage

```
ssm(x, samplingRate = NULL, windowLength = 40, overlap = 75,
  step = NULL, ssmWin = 40, maxFreq = NULL, nBands = NULL,
  MFCC = 2:13, input = c("mfcc", "audiogram", "spectrum")[1],
  norm = FALSE, simil = c("cosine", "cor")[1], returnsSSM = TRUE,
  kernelLen = 200, kernelSD = 0.2, plot = TRUE, specPars = list(levels =
  seq(0, 1, length = 30), color.palette = seewave::spectro.colors, xlab =
  "Time, s", ylab = "kHz", ylim = c(0, maxFreq/1000)), ssmPars = list(levels =
  seq(0, 1, length = 30), color.palette = seewave::spectro.colors, xlab =
  "Time, s", ylab = "Time, s", main = "Self-similarity matrix"),
  noveltyPars = list(type = "b", pch = 16, col = "black", lwd = 3))
```

Arguments

x	path to a .wav file or a vector of amplitudes with specified samplingRate
samplingRate	sampling rate of x (only needed if x is a numeric vector, rather than a .wav file)
windowLength	length of FFT window, ms
overlap	overlap between successive FFT frames, %
step	you can override overlap by specifying FFT step, ms
ssmWin	window for averaging SSM, ms
maxFreq	highest band edge of mel filters, Hz. Defaults to $\text{samplingRate} / 2$. See melfcc
nBands	number of warped spectral bands to use. Defaults to $100 * \text{windowLength} / 20$. See melfcc
MFCC	which mel-frequency cepstral coefficients to use; defaults to 2:13)

input	either MFCCs ("cepstrum") or mel-filtered spectrum ("audiogram")
norm	if TRUE, each FFT frame is normalized by max power
simil	method for comparing frames: "cosine" = cosine similarity, "cor" = Pearson's correlation
returnSSM	if TRUE, returns the SSM
kernelLen	length of checkerboard kernel for calculating novelty, ms (the larger, the more global vs. local the novelty)
kernelSD	SD of checkerboard kernel for calculating novelty
plot	if TRUE, plots the SSM
specPars	graphical parameters passed to <code>seewave::filled.contour.modif2</code> and affecting the spectrogram
ssmPars	graphical parameters passed to <code>seewave::filled.contour.modif2</code> and affecting the plot of SSM
noveltyPars	graphical parameters passed to <code>lines</code> and affecting the novelty contour

Value

If `returnSSM` is TRUE, returns a list of two components: `$ssm` contains the self-similarity matrix, and `$novelty` contains the novelty vector. If `returnSSM` is FALSE, only produces a plot.

References

- El Badawy, D., Marmaroli, P., & Lissek, H. (2013). Audio Novelty-Based Segmentation of Music Concerts. In *Acoustics 2013* (No. EPFL-CONF-190844)
- Foote, J. (1999, October). Visualizing music and audio using self-similarity. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)* (pp. 77-80). ACM.
- Foote, J. (2000). Automatic audio segmentation using a measure of audio novelty. In *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on* (Vol. 1, pp. 452-455). IEEE.

Examples

```
## Not run:
sound = c(soundgen(), soundgen(nSyl = 4, sylLen = 50, pauseLen = 70,
    formants = NA, pitchAnchors = list(time = c(0, 1),
    value = c(500, 330))))
# playme(sound)
m1 = ssm(sound, samplingRate = 16000,
    input = 'audiogram', simil = 'cor', norm = FALSE,
    ssmWin = 10, kernelLen = 150) # detailed, local features
m2 = ssm(sound, samplingRate = 16000,
    input = 'mfcc', simil = 'cosine', norm = TRUE,
    ssmWin = 50, kernelLen = 600) # more global
# plot(m2$novelty, type='b') # use for peak detection, etc

## End(Not run)
```

Index

*Topic **datasets**

- defaults, [13](#)
 - notesDict, [24](#)
 - permittedValues, [27](#)
 - pitchManual, [27](#)
 - presets, [28](#)
 - segmentManual, [32](#)
- addVectors, [2](#)
- analyze, [3](#)
- analyzeFolder, [7](#), [24](#)
- compareSounds, [10](#)
- crossFade, [12](#)
- defaults, [13](#)
- dtw, [11](#), [21](#)
- env, [29](#), [31](#)
- findformants, [4](#), [8](#)
- getRolloff, [14](#), [14](#), [35](#)
- getSigmoid, [15](#)
- getSmoothContour, [16](#)
- getSpectralEnvelope, [18](#), [35](#)
- HzToSemitones, [20](#), [33](#)
- lines, [41](#)
- matchPars, [10](#), [21](#)
- melfcc, [40](#)
- morph, [23](#)
- notesDict, [24](#)
- optim, [5](#), [9](#), [24](#), [25](#)
- optimizePars, [24](#), [30](#)
- permittedValues, [26](#)
- pitchManual, [27](#)
- play, [28](#), [36](#)
- playme, [28](#)
- plot, [30](#), [31](#)
- presets, [28](#)
- segment, [29](#), [32](#)
- segmentFolder, [24](#), [30](#)
- segmentManual, [32](#)
- semitonesToHz, [33](#)
- soundgen, [16](#), [21](#), [23](#), [33](#), [37](#)
- soundgen_app, [37](#)
- spectro, [38](#)
- spectrogram, [6](#), [10](#), [36](#), [38](#)
- ssm, [40](#)