

LIST OF TOOLBOXES

<i>Chapter Introduced</i>	<i>Name of toolbox Folder</i>
Chapter 1	display_speech_1
Chapter 2	analysis
Chapter 4	display_speech_egg
Chapter 6	formant
Chapter 7	vocos and display_speech
Chapter 8	time
Chapter 9	vocal_fold
Chapter 10	formant_track and artm

INDEX

- adaptation, 11, 53
- analysis (*see* linear prediction, spectral analysis, speech analysis)
- articulation, place of, 11, 63, 65, 396–445
- alveolar (gums), 63
 - articulators, 11, 61–62
 - constriction, 11
 - dental (teeth), 63
 - glottal (glottis), 63
 - labial (lips), 63
 - labiodental (lips and teeth), 63
 - palatal (palate), 63
 - velar (soft palate), 63
- articulatory speech synthesis, 396–445
 - acoustic models, 403
 - articulatory models, 398–399
 - cross-sectional area, 400
 - equivalent circuits, 415
 - excitation at the glottis, 409
 - excitation in the vocal tract, 410, 418
 - excitation source models, 409
 - formant frequencies, 401
 - glottal impedance, 411, 418
 - hybrid time-frequency, 420
 - inverse filtering, 424, 430
 - midsagittal distance models, 398
 - model of human vocal system, 422
 - nasal tract, 407, 417
 - noise source models, 413
 - parametric area models, 398
 - radiation models, 408
 - Reynolds number, 415
 - section lengths, 400
 - sentence, optimization, 434
 - simulated annealing algorithm, 427, 429
 - sinus cavities, 407, 417
 - sound propagation, 404
 - subglottal models, 411
 - subglottal system, 418
 - synthesis of affricate consonants, 436
 - synthesis of stop consonants, 436
 - synthesis of turbulence, 440
 - synthesis of unvoiced consonants, 435
 - synthesis of unvoiced speech, 434–435
 - transfer function of affricate /CH/, 441
 - transmission line model, 406
 - turbulence noise models, 438
 - turbulence, 414–415
 - uniform lossless section, 405
 - velopharyngeal port, 417–418
 - velum, 418
 - vocal tract, 400
 - vocal tract area, 425
 - vocal tract models, 403
 - vocal tract sections, number of, 416
 - vowels, optimization, 433
 - wall impedance, 406
 - wave digital filter, 419
 - wave propagation, 405
- articulatory speech synthesis (artsm)
 - toolbox, 211, 215
 - acoustic model, 213
 - animation, 241
 - articulatory positions settings, 219
 - articulatory vocal tract model, 220
 - aspiration noise, 228
 - excitation option, 225
 - excitation parameter, 227
 - excitation waveform, 231
 - excitation within vocal tract, 232
 - jitter/shimmer parameters, 228
 - load/mark option, 216
 - metropolis algorithm, 222
 - mixed excitation, 232
 - optimization option, 220
 - shape option, 219
 - simulated annealing, 222
 - steps in articulatory speech synthesis, 213
 - subglottal parameters, 229
 - synthesis option, 239
 - synthesizer, 212
 - target formants, 214
 - turbulence excitation models, 232
 - turbulence noise source, 233
 - unvoiced excitation, 231
 - unvoiced (turbulence) excitation models, 233
 - voiced excitation, 226
 - voiced excitation model, 228
 - voiced parameters window, 227
- ASCII data file, 19, 24–28
- ASCII to wav, 24–25
- aspiration (*see* sound classification)
- assimilation, 11, 53–54
- autocorrelation (*see* linear prediction and speech analysis toolbox)
- cepstrum, 45, 295
- coarticulation, 11, 53
- consonants (*see* sound classifications)
- contours, 15
 - aspiration, 15
 - duration, 53
 - energy (*see* speech analysis toolbox)
 - formant, 15, 43–44
 - frication, 15
 - gain, 15, 168–169
 - jitter, 44
 - nasal, 15
 - pitch, 15, 43–44, 52–53, 68–69, 166–168
 - silence, 15, 48
 - voiced/unvoiced/mixed excitation sounds, 15
- data, 267–289
- data windowing (*see* windows)
- display_speech toolbox, 174
- display_speech_1 toolbox, 13
- display_speech_egg toolbox, 88
- electroglottograph, 16–17, 86
- energy, 290–291
 - average magnitude, 291
 - short-time average, 290
 - zero-crossing, 291
- excitation, 308–319
 - aspiration noise, 314
 - glottal models, 308
 - LF model, 308
 - mixed, 59
 - noise model, 314

- excitation (*continued*)
 polynomial model, 318
 turbulent noise model, 314
 unvoiced, 59, 65
 voiced, 59, 65, 318
 volume-velocity, 17, 80
- file attribute conversion, 1–2
- filters, 296
 comb, 296
 litters, 296
 long-pass, 296
 notch, 296
 short-pass, 296
- formants, 15
 formant contours, 15
 formant frequencies (formants), 15, 60, 68
 natural frequencies (tube closed), 61
 resonances of vocal tract, 15, 72
- formant synthesizer toolbox, 139
- formant track toolbox, 214
- frication (*see* sound classifications)
- fundamental frequency of voicing, 15, 57, 86
 (*see* speech, pitch)
- glottis, 55, 57, 63, 80 (*see* laryngeal function)
 glottal area, 57
 glottal closure, 79, 313
 glottal excitation, 61
 glottal features (*see* volume-velocity)
 glottal flow (*see* volume-velocity)
 glottal inverse filtering, 39, 41
 glottal pulse, 59, 61, 313
 glottal pulse skewness, 313
 glottal pulse width, 313
 glottal volume velocity, 41, 59 (*see* volume-velocity)
- graphic user interface (GUI), 6
- harmonics richness factor (HRF), 313
- harmonics-to-noise ratio (HNR), 313
- intelligibility of speech, 17, 473
- inverse filtering, 15, 40, 81–82 (*see* linear prediction and speech)
 deconvolution, 15
- labeling of sounds (*see* sound classifications)
 allophones, 63
 ARPAbet, 63–64
 classification, 11
 demisyllables, 11
 diphones, 11
 dyads, 11
- International Phonetic Alphabet (IPA), 63
- key word, 63
- morphs, 11
- phonetics, 11
- phonetic alphabet, 63
- phonemic transcription, 53, 64
- language, 51
- laryngeal function, 54, 75
 16-mass model, 380
 abducted, 59
 adducted, 59
 acoustic model, 385
 Bernoulli principle, 378
 body-cover model (three-mass model), 382
 cinematography, 76–77–78
 contact area, 394
 continuum model, 381
 data, 75, 267–292
 direct/indirect laryngoscopy, 75, 77, 80
 electroglottographic waveform, 81
 electroglottography, 80–81
 equivalent circuit, 386
 fiberoptics, 77
 glottal area, 378
 glottal closure, 79
 glottis, 386
 glottographic waveforms, 393
 glottography (optical/electrical), 79–80
 laryngeal evaluative methods, 76
 laryngeal microscopy, 76
 laryngograph, 80
 laryngoscope, 75, 78
 larynx, 54, 80, 369
 magnetic resonance imaging, 79
 measurements, 75
 mechanical models, 379
 membranes, 372
 multiple-mass model, 380
 muscles, 372
 myoelastic-aerodynamic theory, 377
 observational, 75
 one-mass model, 379
 projected glottal area, 394
 radiography, 79
 ribbon model, 381, 390–393, 395
 simulation of EGG waveform, 395
 skeletal structure of the larynx, 369
 still photography, 76
 stroboscope, 77
 stroboscopic laryngoscopy, 77
 tissue layers, 376
 two-mass vocal fold model, 379, 383–389
 ultrasound, 79
 ventricular folds, 375
- vibration theory, 377–378, 393
- vocal fold, 77, 393
- vocal fold vibratory models, 368–395
- vocal tract, 385
- volume-velocity, 80
- x-rays of, 79
- linear prediction (LP), 16, 35–37, 81, 95
 Akaike information criterion (AIC), 118
 algorithms, 121
 appending zeros, 105
 autocorrelation, 97, 99
 autocorrelation matrix, 102
 autocorrelation method, 113
 autoregression (AR), 109, 119
 autoregression-moving average (ARMA), 107, 109, 118–119
 Blackman-Tukey spectral estimate, 105
 Burg's spectral estimator, 122
 Cholesky decomposition, 116, 122
 classical estimation, 103
 correlation matrix, 116
 covariance method, 35, 115
 date adaptive methods, 107
 Durbin's algorithm, 118
 error criterion, 36
 estimates of power spectrum, 103
 formant bandwidth, 123
 formant frequency, 123
 glottal excited linear prediction (GELP), 319
 inverse filter model, 113
 lags, number of, 100
 least squares estimation, 111
 Levinson algorithm, 115, 121
 linear prediction equations, 116
 linear prediction model, 113
 linear predictive coding (LPC), 118
 log area coefficients, 124
 LP power spectral estimate, 117
 maximum entropy estimator, 107
 moving average (MA), 109, 118–119
 normal equations, 112, 114–115, 121
 order selection, 36
 partial correlation coefficients (PARCOR), 122, 124
 perceptual linear prediction, 36
 periodogram, 103
 positive definite, 102
 power spectral estimation/linear prediction, 116
 prediction error, 114–115
 prediction power, 114
 pre-emphasis, 123

- principal of orthogonality, 112
 recursive maximum likelihood, 36
 reflection coefficient, 122
 segmenting the autocorrelation
 function, 105
 sidelobes, 105
 spectral balance, 123
 spectral flatness, 118
 spectral resolution, 105
 statistical stability, 105
 Toeplitz matrix, 102, 121
 unbiased autocorrelation estimate,
 97
 vocal tract model, 96
 window frame length, 35
 Wiener-Hopf equations, 114
 Yule-Walker equations, 114
- linguistics, 51–52
 grammar, 52
 semantics, 52
 sentences, 52
- matlab, 1, 13, 19, 24–25
 environment, 8
 error messages, 5–6
 graphic user interface 4, 6
 M-files, 1, 3
 P-files, 1, 3
 regular version, 6–7
 runtime server, 1–6
- maximum phase, 297
 microphone, 82–84
 minimum phase, 297
- models
 acoustic tube, 12
 filter, 12
 glottal area, 17
 glottal excitation, 308
 inverse filtering, 40, 43
 LF, 308
 oral-nasal tracts, 12, 65
 pole-zero, 12
 production, 12
 sound, 62
 source-filter, 62
 spectra, 12
 speech production, 56
 speech radiation, 12
 vocal tract, 40–42
 waveform, 12
- myeloelastic-aerodynamic theory, 57,
 377
 Bernoulli effect, 57, 377
 subglottal pressure, 57
- noise energy, normalized (NNE), 313
- open quotient, 313
 operating system and platform, 7
- regular Matlab version, 6–7
 runtime version, 2–6
- orthography, 53
- perturbation (*see* speech, jitter,
 shimmer)
- pharynx (*see* vocal tract and laryngeal
 function)
- phonemes (*see* sound classifications)
- phonetics, 51
- phonemic transcription, 53
- pitch (*see* fundamental frequency of
 voicing and speech)
- pneumotachographic mask, 81
- quality of speech, 17, 473
- radiation, acoustic model, 389
- residue, 313
- Rothenburg mask (*see*
 pneumotachographic mask)
- sampling frequency, 23
- signal-to-noise, 313
- silence, 15
- software, 1–7
 installation, 1–7
 interactive, 16
- sound board, 13, 22–23, 29
- sound classifications, 11–12, 51
 acoustic, 63
 affricates, 65
 allophones, 67
 articulation, 51, 63, 65–66
 aspiration, 15, 65
 aspiration noise, 313
 consonants, 11, 52–53, 59, 61, 63,
 65, 67
 continuants, 65
 diphthongs, 64–65
 fricatives, 11, 59, 65–66
 intensity, 52
 jitter/shimmer, 65
 nasal, 11, 65–66, 72
 noncontinuants, 66
 nonsibilants, 66
 obstruents, 66
 phone, 51
 phonemes, 11, 15–16, 51–52, 54,
 62–63, 67, 70–72
 pitch, 52, 69
 plosives, 11, 59, 65–66
 schwa, 64
 semivowels/liquids/glides, 65–66
 sibilants, 11, 66
 sonorants, 66
 spectral characteristics, 67
 statistics, 67
 stops, 11, 59, 65–66
- suprasegmental, 53
- tense/lax, 64
- timing, 52
- turbulent noise, 72
- voice bar, 67
- voiced/unvoiced/mixed excitation,
 11, 15, 59, 63 65
- vowels, (*see* vowel)
- whisper, 67
- sound pressure levels, 83
- spectra (*see* spectral analysis)
- spectral analysis, 34, 136
 algorithms, 15, 35, 38
 Akaike information criterion (AIC),
 36
 Akaike maximum likelihood
 estimate (Akaike MLE), 38
 autoregression-moving average
 (ARMA), 37–39
 autoregression (AR), 15
 criterion autoregressive transfer
 function (CAT), 36
 data, 31, 267–289, 308–319
 error criterion, 36
 esprit, 34
 FFT, 36–37, 39
 final prediction error (FPE), 36
 linear prediction (LP), 15, 36–37
 Mayne-Firoozan method, 38
 modified Yule-Walker likelihood
 estimate (MYLE), 38
 moving average, 15, 38
 moving average (MA), 38, 40
 music, 34
 order selection, 36–37
 perceptual linear prediction, 36
 periodogram, 34
 poles, number of, 35
 pole-zero plot, 40
 recursive maximum likelihood, 36
 spectra, 17, 34, 275
 spectral estimation, 34–35
 spectral tilt, 313
 window frame length, 35
 word boundary identification, 72
 Yule-Walker, 114, 292
- spectral estimation (*see* spectral
 analysis)
- spectrogram, 13–14, 17, 32–34, 83–85
- speech
 accents, 62
 acoustic tube (*see* vocal tract)
 analysis, 8, 16, 19
 analysis-by-synthesis, 134
 articulatory movements, 62
 articulatory speech synthesis (*see*
 articulatory speech synthesis)
 assessing synthetic speech, 474
 autocorrelation function, 15

- speech (*continued*)
- characteristics, 11
 - characteristics of voice, 136, 308–319
 - classification, 300, 304 (*see* classification of sounds)
 - coding, 16
 - data, 267–289
 - dialects, 62
 - EGG, 267–289
 - electroglottographic signals, 16–17
 - energy, 15
 - excitation, 11, 59, 72
 - features, 15–16
 - formant tracking, 294
 - fundamental frequency, 59–60 (*see* speech, pitch)
 - glottal area, 267
 - glottal flow, 86
 - human production, 11–12
 - human vocal system, 55
 - intelligibility, 15, 60, 473, 475
 - intensity, 86
 - jitter, 85
 - listening tests, 476
 - modeling, 16
 - models, 300
 - oral-nasal tracts, 65
 - perception, 12
 - phonemes, 15–16, 54, 62
 - pitch asynchronous analysis, 15, 40, 44
 - pitch perturbation, 294–295
 - pitch synchronous analysis, 15, 40
 - pitch tracking, 294–295
 - pole-zero, 37
 - production, 16, 54
 - prosody, 52
 - quality, 15, 61–62, 473, 475
 - radiation, 12
 - recognition, 8, 10
 - Russian vowels, 275
 - segmenting-labeling (*see* labeling)
 - semantics, 52
 - shimmer, 85
 - source excitation waveform, 61
 - speaking rate, 62
 - speaking styles, 62
 - spectra (*see* spectral analysis)
 - spectrograms, 275
 - speech analysis toolbox (*see* speech analysis toolbox)
 - syllables, 51
 - synthesis, 8–9, 16, 53, 62
 - synthesis systems, 15
 - synthetic speech quality, 477
 - time modification, 17 (*see* time modification)
 - toolbox, 19
 - vocal tract area, 275
 - voice, 313, 320
 - voice (vocal) quality, 60, 86, 473
 - voiced or unvoiced, 15, 48, 59, 65, 72
 - volume-velocity, 17, 80–82
 - waveforms, 275
 - word, 52
 - zero crossing rate, 15
- speech analysis (*see* speech, spectral analysis, and time domain analysis)
- speech analysis toolbox, 19–50
- ARMA analysis, 37
 - asynchronous analysis, 40, 44
 - autocorrelation, 30–32, 35, 291–292
 - cepstrum, 45
 - cross hair, 13, 26–29
 - cursor, 26–27, 29
 - cut, 25, 27–28
 - dialects, 61
 - edit, 25–29
 - energy-ZCR option, 29–31
 - esprit, 34
 - FFT, 34
 - file menu and window, 19, 22, 26
 - frequency-domain analysis, 32–34 (*see* spectral analysis)
 - frequency response (transfer function), 40–41
 - formant, 44
 - input signal window, 19, 22, 27
 - insert, 25, 27–28
 - jitter, 44
 - LP analysis, 35, 95
 - MA analysis, 38
 - main menu window, 22
 - pitch, 44
 - play, 28–29
 - property window, 30–31, 33, 35, 40
 - rand, 25
 - randn, 25
 - random waveform generation, 25
 - rate conversion, 19, 22–23
 - spectra, 34
 - spectrogram window, 33–34
 - time domain analysis, 29
 - waveform generator, 19, 25
 - window length, 30
 - window of data, 31
 - window option, 30
 - WRLS-VFF, 46
 - ZCR calculation, 32
- speech synthesis, 53, 72, 128, 136
- analysis-by-synthesis, 134
 - articulatory, 17, 128
 - characteristics of voice, 136
 - excitation source, 137
 - formant synthesis, 16, 41–42, 72, 128, 130
 - glottal excited, linear prediction (GELP), 319
 - intonation, 53
 - linear prediction, 128
 - LP synthesis, 131
 - pitch synchronous overlap-add (PSOLA), 133
 - rate of speaking, 61
 - recognition, 72
 - sinusoidal, 128
 - stress, 53
 - text to speech, 9, 128, 133
 - vocoder, 128
 - word-sentences, 28
- speed quotient, 313
- standards, 265
- subglottal pressure, 59
- supraglottal pressure, 59
- synthetic speech, 17, 128, 152, 179, 211 (*see* speech synthesis)
- text to speech, 9
- time domain analysis (*see* speech analysis toolbox)
- time modification of speech, 330–367
- boundary detection, 353
 - errors, 358
 - feature detection, 335
 - feature scores, 338
 - formant tracking, 347
 - frame doubling, 360
 - frame skipping, 360
 - fricative detection, 350
 - glitch prevention, 366
 - labeling, 330, 356
 - linear prediction coding (LPC) synthesizer, 360
 - mapping, 362
 - nasal detection, 347
 - sampling method, 331
 - segment detection, 334
 - segment labeling, 356
 - segmentation, 330
 - semivowel detection, 348
 - sonorant detection, 342
 - speech segment, 330
 - time modification and synthesis, 330, 364
 - unvoiced stop detection, 351
 - v/u/s boundary detection, 354
 - variable-playback-rate method, 331
 - vocoder method, 332
 - voice bar detection, 345
 - voiced consonant detection, 343
 - voiced fricative detection, 350
 - volume function, 337
 - vowel detection, 342

- time modification of speech toolbox, 179
 analysis option, 182
 main function window, 182
 modify option, 184
 time modification/synthesis, 189
- vocal_fold toolbox, 201
 ribbon vocal fold model, 206
 two-mass vocal fold model, 202
 vocal fold models, 201
- vocal folds, 54, 56–58, 60, 77, 80 (*see* laryngeal function)
 abruptness of closure, 313
 elastic, 57
 excitation source, 449
 myoelastic-aerodynamic theory of, 368
 tense, 57
 ultra high-speed film, 58
 vocal fold vibratory motion, 58
 vocal folds or cords, 54, 56–58, 60, 77
 ventricular (false vocal), 54
 volume-velocity, 80
- vocal system, 54, 58 (*see* laryngeal function, vocal folds, and vocal tract)
 cricoid cartilage, 58
 epiglottis, 58
 false vocal fold, 58
 thyroarytenoid muscle, 58
 thyroid cartilage, 58
 trachea, 58
 true vocal fold, 58
 ventricle, 58
 volume-velocity, 58, 60
- vocal tract, 11–12, 17, 51, 53, 58, 60–61, 66, 72
 acoustic equations, 453
 acoustic transfer function, 447
 constrictions, 72
 equivalent circuit, 387–388
 excitation, 11, 15
 excitation source, 449
 excitation source located in vocal tract, 450–451
 mouth, 60, 65
 nasal cavity (tract), 60, 65, 72
 oral cavity (tract), 51, 72
 pharynx, 51, 60
 resonant cavities, 60, 72
 schematic of, 11–12
 soft palate, 60
 thermal conduction, 387–388
 tongue, 60
 velum, 11, 60, 65, 72, 418
 viscous friction, 387–388
 wall, 386–387
- voice, characteristics of, 61–62, 308–319
- voice conversion, 9, 17, 320
- voice conversion (vocos) toolbox, 152
 correction, 163
 gain contour, 168
 glottal waveform, 169
 linear prediction (LP), 152
 modification, 164
 pitch contour, 166
 spectrogram (formants), 170
 speech analysis-formant, 162
 synthesize, 160
 voice conversion, 171
 voice conversion-segmentation, 173
 voice type, 165
- voice modification and synthesis, 320–329
 affine model, 327
 gain contour modification, 324
 glottal closure instants (GCIs), 320
 glottal pulse modification, 326
 intensity, 324
 jitter, 323
 loudness, 324
 peak picking, 321
 pitch contour model, 322
 pitch contour modification, 322–323
 pitch detection, 320
 pitch period estimation, 320
 pitch wave, 322–324
 shimmer, 324
 speaker translation models, 326
 synthesis, 323, 325
 vocal tract modification, 326
 voice conversion models, 326
 voice conversion, 326
- voice types, 308–319
 breathy, 15, 313
 falsetto, 60, 313
 female, 15
 harsh, 15, 313
 hoarse, 15, 313
 male, 15
 modal (normal), 60, 313
 vocal fry (creaky), 60, 313
- volume-velocity, 308–319
 abruptness of closure, 313
 aspiration noise, 313
 breathy, 313
 glottal closure, 313
 glottal factors, 313
 glottal pulse skewness, 313
 glottal pulse width, 313
 harmonics richness factor (HRF), 313
 harmonics-to-noise ratio, 313
 hoarse, 313
 jitter, 313
- modal, 313
 normalized noise energy (NNE), 313
 open quotient, 313
 pitch period, 313
 pitch skewness, 313
 pulse width, 313
 residue, 313
 rough, 313
 shimmer, 313
 signal-to-noise ratio, 313
 spectral tilt, 313
 speed quotient, 313
 vocal fry, 313
 voice, 313
- volume-velocity waveform (*see* speech, vocal folds, laryngeal function, excitation)
- vowel, 11, 17, 53, 55, 59, 62–64, 67–68, 70
 articulatory speech synthesizer, 68
 bandwidths, 68, 70
 cross-sectional vocal tract area, 67
 features, 446
 formant frequencies, 68–70
 phoneme labeling, 68
 pitch contours, 68–69
 spectrogram, 67, 69
 spectrum, 67
 triangle, 71
 vocal tract shape, 67
 waveform, 67, 69
- wav to ASCII, 23–24
- waveforms, 17, 43 (*see* speech and volume-velocity)
 excitation, 62
 zero-crossings, 72
- wave propagation, 95
 equation of continuity, 95
 equation of motion, 95
- window length, overlap, 29
- window of data, 31
- windows
 bartlett, 30
 blackman, 30
 blackman-tukey, 34
 burg, 35
 chebyshev, 30
 hamming, 30
 hanning, 30
 kaiser, 30
 rectangular, 30
 triangular, 30
- word boundaries, 16
- zero-crossing (ZCR), 32

