

CPC**COOPERATIVE PATENT CLASSIFICATION****G10L****SPEECH ANALYSIS OR SYNTHESIS; SPEECH RECOGNITION;
SPEECH OR VOICE PROCESSING; SPEECH OR AUDIO CODING OR
DECODING****NOTE**

This subclass does not cover:

devices for the storage of speech signals, which are covered by subclasses [G11B](#) and [G11C](#) ;

encoding of compressed speech signals for transmission or storage, which is covered by group [H03M 7/30](#).

G10L 13/00**Speech synthesis; Text to speech systems**

- G10L 13/02 . Methods for producing synthetic speech; Speech synthesisers
- G10L 2013/021 .. { **Overlap-add techniques** }
- G10L 13/027 .. Concept to speech synthesisers; Generation of natural phrases from machine-based concepts (**generation of parameters for speech synthesis out of text [G10L 13/08](#)**)
- G10L 13/033 .. Voice editing, e.g. manipulating the voice of the synthesiser
- G10L 13/0335 ... { **Pitch control** }
- G10L 13/04 .. Details of speech synthesis systems, e.g. synthesiser structure or memory management
- G10L 13/043 ... { **Synthesisers specially adapted to particular applications** }

WARNING

This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L 13/00](#) and subgroups.

- G10L 13/047 ... Architecture of speech synthesisers
- G10L 13/06 . Elementary speech units used in speech synthesisers; Concatenation rules
- G10L 13/07 .. Concatenation rules
- G10L 13/08 . Text analysis or generation of parameters for speech synthesis out of text, e.g. grapheme to phoneme translation, prosody generation or stress or intonation determination
- G10L 2013/083 .. { **Special characters, e.g. punctuation marks** }
- G10L 13/086 .. { **Detection of language** }
- G10L 13/10 .. Prosody rules derived from text; Stress or intonation
- G10L 2013/105 ... { **Duration** }

G10L 15/00**Speech recognition ([G10L 17/00](#) takes precedence)**

- G10L 15/005 . {Language recognition }
- G10L 15/01 . Assessment or evaluation of speech recognition systems
- G10L 15/02 . Feature extraction for speech recognition; Selection of recognition unit
- G10L 2015/022 .. { Demisyllables, biphones or triphones being the recognition units }
- G10L 2015/025 .. { Phonemes, fenemes or fenones being the recognition units }
- G10L 2015/027 .. { Syllables being the recognition units }
- G10L 15/04 . Segmentation; Word boundary detection
- G10L 15/05 .. Word boundary detection
- G10L 15/06 . Creation of reference templates ; Training of speech recognition systems, e.g. adaptation to the characteristics of the speaker's voice ([G10L 15/14 takes precedence](#))
- G10L 15/063 .. { Training }
- G10L 2015/0631 ... { Creating reference templates; Clustering }
- G10L 2015/0633 { using lexical or orthographic knowledge sources }
- G10L 2015/0635 ... { updating or merging of old and new templates; Mean values; Weighting }
- G10L 2015/0636 { Threshold criteria for the updating }
- G10L 2015/0638 ... { Interactive procedures }
- G10L 15/065 .. Adaptation
- G10L 15/07 ... to the speaker
- G10L 15/075 { supervised, i.e. under machine guidance }
- G10L 15/08 . Speech classification or search
- G10L 2015/081 .. { Search algorithms, e.g. Baum-Welch or Viterbi }
- G10L 15/083 .. {Recognition networks ([G10L 15/142](#), [G10L 15/16 take precedence](#)) }
- G10L 2015/085 .. { Methods for reducing search complexity, pruning }
- G10L 2015/086 .. { Recognition of spelled words }
- G10L 2015/088 .. { Word spotting }
- G10L 15/10 .. using distance or distortion measures between unknown speech and reference templates
- G10L 15/12 .. using dynamic programming techniques, e.g. dynamic time warping [DTW]
- G10L 15/14 .. using statistical models, e.g. hidden Markov models [HMMs] ([G10L 15/18 takes precedence](#))
- G10L 15/142 ... { Hidden Markov Models [HMMs] }
- G10L 15/144 {Training of HMMs }
- G10L 15/146 {with insufficient amount of training data, e.g. state sharing, tying, deleted interpolation }
- G10L 15/148 {Duration modelling in HMMs, e.g. semi HMM, segmental models or transition probabilities }
- G10L 15/16 .. using artificial neural networks
- G10L 15/18 .. using natural language modelling
- G10L 15/1807 ... { using prosody or stress }

- G10L 15/1815 . . . { Semantic context, e.g. disambiguation of the recognition hypotheses based on word meaning }
- G10L 15/1822 . . . { Parsing for meaning understanding }
- G10L 15/183 . . . using context dependencies, e.g. language models
- G10L 15/187 Phonemic context, e.g. pronunciation rules, phonotactical constraints or phoneme n-grams
- G10L 15/19 Grammatical context, e.g. disambiguation of the recognition hypotheses based on word sequence rules
- G10L 15/193 Formal grammars, e.g. finite state automata, context free grammars or word networks
- G10L 15/197 Probabilistic grammars, e.g. word n-grams
- G10L 15/20 . . . Speech recognition techniques specially adapted for robustness in adverse environments, e.g. in noise, of stress induced speech ([G10L 21/02 takes precedence](#))
- G10L 15/22 . . . Procedures used during a speech recognition process, e.g. man-machine dialogue
- G10L 2015/221 . . . Announcement of recognition results
- G10L 15/222 . . . { Barge in, i.e. overridable guidance for interrupting prompts }
- G10L 2015/223 . . . Execution procedure of a spoken command
- G10L 2015/225 . . . Feedback of the input speech
- G10L 2015/226 . . . Taking into account non-speech characteristics
- G10L 2015/227 of the speaker; Human-factor methodology
- G10L 2015/228 of application context
- G10L 15/24 . . . Speech recognition using non-acoustical features
- G10L 15/25 . . . using position of the lips, movement of the lips or face analysis
- G10L 15/26 . . . Speech to text systems ([G10L 15/08 takes precedence](#))
- G10L 15/265 . . . { Speech recognisers specially adapted for particular applications (devices for signalling identity of wanted subscriber in a telephonic communication equipment controlled by voice recognition [H04M 1/271](#); speech interaction details in interactive information services in a telephonic communication system [H04M 3/4936](#)) }

WARNING

This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L 15/00](#) and subgroups.

- G10L 15/28 . . . Constructional details of speech recognition systems
- G10L 15/285 . . . {Memory allocation or algorithm optimisation to reduce hardware requirements }
- G10L 15/30 . . . Distributed recognition, e.g. in client-server systems, for mobile phones or network applications
- G10L 15/32 . . . Multiple recognisers used in sequence or in parallel; Score combination systems therefor, e.g. voting systems
- G10L 15/34 . . . Adaptation of a single recogniser for parallel processing, e.g. by use of multiple processors or cloud computing

G10L 17/00 Speaker identification or verification

- G10L 17/005 . { Speaker recognisers specially adapted for particular applications ([G07C 9/00071](#) takes precedence) }
- WARNING**
- This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L 17/00](#) and subgroups.
- G10L 17/02 . Preprocessing operations, e.g. segment selection; Pattern representation or modelling, e.g. based on linear discriminant analysis [LDA] or principal components; Feature selection or extraction
- G10L 17/04 . Training, enrolment or model building
- G10L 17/06 . Decision making techniques; Pattern matching strategies
- G10L 17/08 . . Use of distortion metrics or a particular distance between probe pattern and reference templates
- G10L 17/10 . . Multimodal systems, i.e. based on the integration of multiple recognition engines or fusion of expert systems
- G10L 17/12 . . Score normalisation
- G10L 17/14 . . Use of phonemic categorisation or speech recognition prior to speaker recognition or verification
- G10L 17/16 . Hidden Markov models [HMMs]
- G10L 17/18 . Artificial neural networks; Connectionist approaches
- G10L 17/20 . Pattern transformations or operations aimed at increasing system robustness, e.g. against channel noise or different working conditions
- G10L 17/22 . Interactive procedures; Man-machine interfaces
- G10L 17/24 . . the user being prompted to utter a password or a predefined phrase
- G10L 17/26 . Recognition of special voice characteristics, e.g. for use in lie detectors; Recognition of animal voices
- G10L 19/00** **Speech or audio signal analysis-synthesis techniques for redundancy reduction, e.g. in vocoders; Coding or decoding of speech or audio signal, using source filter models or psychoacoustic analysis (in musical instruments [G10H](#))**
- G10L 2019/0001 . { Codebooks }
- G10L 2019/0002 . . { Codebook adaptations }
- G10L 2019/0003 . . { Backward prediction of gain }
- G10L 2019/0004 . . { Design or structure of the codebook }
- G10L 2019/0005 . . . { Multi-stage vector quantisation }
- G10L 2019/0006 . . . { Tree or treillis structures; Delayed decisions }
- G10L 2019/0007 . . { Codebook element generation }
- G10L 2019/0008 . . . { Algebraic codebooks }

- G10L 2019/0009 . . . { Orthogonal codebooks }
- G10L 2019/001 . . . { Interpolation of codebook vectors }
- G10L 2019/0011 . . { Long term prediction filters, i.e. pitch estimation }
- G10L 2019/0012 . . { Smoothing of parameters of the decoder interpolation }
- G10L 2019/0013 . . { Codebook search algorithms }
- G10L 2019/0014 . . . { Selection criteria for distances }
- G10L 2019/0015 . . . { Viterbi algorithms }
- G10L 2019/0016 . . { Codebook for LPC parameters }

- G10L 19/0017 . { Lossless audio signal coding; Perfect reconstruction of coded audio signal by transmission of coding error ([G10L 19/24 takes precedence](#)) }

- G10L 19/0018 . {Speech coding using phonetic or linguistic decoding of the source; Reconstruction using text-to-speech synthesis }

- G10L 19/0019 . { Vocoders specially adapted for particular applications }

WARNING

This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L 19/00](#) and subgroups.

- G10L 19/002 . Dynamic bit allocation ([for perceptual audio coders G10L 19/032](#))

- G10L 19/005 . Correction of errors induced by the transmission channel, if related to the coding algorithm

- G10L 19/008 . Multichannel audio signal coding or decoding, i.e. using interchannel correlation to reduce redundancies, e.g. joint-stereo, intensity-coding, matrixing ([arrangements for reproducing spatial sound H04R 5/00](#); stereophonic systems, e.g. spatial sound capture or matrixing of audio signals in the decoded state [H04S](#))

- G10L 19/012 . Comfort noise or silence coding

- G10L 19/018 . Audio watermarking, i.e. embedding inaudible data in the audio signal

- G10L 19/02 . using spectral analysis, e.g. transform vocoders or subband vocoders
- G10L 19/0204 . . { using subband decomposition }
- G10L 19/0208 . . . {Subband vocoders }
- G10L 19/0212 . . { using orthogonal transformation }
- G10L 19/0216 . . . {using wavelet decomposition }
- G10L 19/022 . . Blocking, i.e. grouping of samples in time; Choice of analysis windows; Overlap factoring
- G10L 19/025 . . . Detection of transients or attacks for time/frequency resolution switching
- G10L 19/028 . . Noise substitution, i.e. substituting non-tonal spectral components by noisy source ([comfort noise for discontinuous speech transmission G10L 19/012](#))

- G10L 19/03 . . Spectral prediction for preventing pre-echo; Temporary noise shaping [TNS], e.g. in MPEG2 or MPEG4

- G10L 19/032 . . . Quantisation or dequantisation of spectral components
- G10L 19/035 Scalar quantisation
- G10L 19/038 Vector quantisation, e.g. TwinVQ audio
- G10L 19/04 . . using predictive techniques
- G10L 19/06 . . . Determination or coding of the spectral characteristics, e.g. of the short-term prediction coefficients
- G10L 19/07 Line spectrum pair [LSP] vocoders
- G10L 19/08 . . . Determination or coding of the excitation function ; Determination or coding of the long-term prediction parameters
- G10L 19/083 the excitation function being an excitation gain ([G10L 25/90 takes precedence](#))
- G10L 19/087 using mixed excitation models, e.g. MELP, MBE, split band LPC or HVXC
- G10L 19/09 Long term prediction, i.e. removing periodical redundancies, e.g. by using adaptive codebook or pitch predictor
- G10L 19/093 using sinusoidal excitation models
- G10L 19/097 using prototype waveform decomposition or prototype waveform interpolative [PWI] coders
- G10L 19/10 the excitation function being a multipulse excitation
- G10L 19/107 Sparse pulse excitation, e.g. by using algebraic codebook
- G10L 19/113 Regular pulse excitation
- G10L 19/12 the excitation function being a code excitation, e.g. in code excited linear prediction [CELP] vocoders
- G10L 19/125 Pitch excitation, e.g. pitch synchronous innovation CELP [PSI-CELP]
- G10L 19/13 Residual excited linear prediction [RELP]
- G10L 19/135 Vector sum excited linear prediction [VSELP]
- G10L 19/16 . . . Vocoder architecture
- G10L 19/167 { Audio streaming, i.e. formatting and decoding of an encoded audio signal representation into a data stream for transmission or storage purposes }
- G10L 19/173 { Transcoding, i.e. converting between two coded representations avoiding cascaded coding-decoding }
- G10L 19/18 Vcoders using multiple modes
- G10L 19/20 using sound class specific coding, hybrid encoders or object based coding
- G10L 19/22 Mode decision, i.e. based on audio signal content versus external parameters
- G10L 19/24 Variable rate codecs, e.g. for generating different qualities using a scalable representation such as hierarchical encoding or layered encoding
- G10L 19/26 . . . Pre-filtering or post-filtering
- G10L 19/265 { Pre-filtering, e.g. high frequency emphasis prior to encoding }
- G10L 21/00** **Processing of the speech or voice signal to produce another audible or non-audible signal, e.g. visual or tactile, in order to modify its quality or its intelligibility ([G10L 19/00 takes precedence](#))**
- G10L 21/003 . . . Changing voice quality, e.g. pitch or formants
- G10L 21/007 . . . characterised by the process used
- G10L 21/01 Correction of time axis

- G10L 21/013 ... Adapting to target pitch
- G10L 2021/0135 { Voice conversion or morphing }
- G10L 21/02 . Speech enhancement, e.g. noise reduction or echo cancellation ([reducing echo effects in line transmission systems H04B 3/20](#) ; [echo suppression in hands-free telephones H04M 9/08](#))
- G10L 21/0202 .. { Applications }

WARNING

This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L 21/00](#) and subgroups.

- G10L 21/0205 ... { Enhancement of intelligibility of clean or coded speech }

WARNING

This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L 21/0364](#), [G10L 21/057](#).

- G10L 21/0208 .. Noise filtering
- G10L 2021/02082 ... { the noise being echo, reverberation of the speech }
- G10L 2021/02085 ... { Periodic noise }
- G10L 2021/02087 ... { the noise being separate speech, e.g. cocktail party }
- G10L 21/0216 ... characterised by the method used for estimating noise
- G10L 2021/02161 { Number of inputs available containing the signal or the noise to be suppressed }
- G10L 2021/02163 { Only one microphone }
- G10L 2021/02165 { Two microphones, one receiving mainly the noise signal and the other one mainly the speech signal }
- G10L 2021/02166 { Microphone arrays; Beamforming }
- G10L 2021/02168 { the estimation exclusively taking place during speech pauses }
- G10L 21/0224 Processing in the time domain
- G10L 21/0232 Processing in the frequency domain
- G10L 21/0264 ... characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques
- G10L 21/0272 .. Voice signal separating
- G10L 21/028 ... using properties of sound source
- G10L 21/0308 ... characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques
- G10L 21/0316 .. by changing the amplitude
- G10L 21/0324 ... Details of processing therefor
- G10L 21/0332 involving modification of waveforms
- G10L 21/034 Automatic adjustment
- G10L 21/0356 ... for synchronising with other signals, e.g. video signals
- G10L 21/0364 ... for improving intelligibility

G10L 2021/03643	{ Diver speech }
G10L 2021/03646	{ Stress or Lombard effect }
G10L 21/038	..	using band spreading techniques
G10L 21/0388	...	Details of processing therefor
G10L 21/04	.	Time compression or expansion
G10L 21/043	..	by changing speed
G10L 21/045	...	using thinning out or insertion of a waveform
G10L 21/047	characterised by the type of waveform to be thinned out or inserted
G10L 21/049	characterised by the interconnection of waveforms
G10L 21/055	..	for synchronising with other signals, e.g. video signals
G10L 21/057	..	for improving intelligibility
G10L 2021/0575	...	{ Aids for the handicapped in speaking }
G10L 21/06	.	Transformation of speech into a non-audible representation, e.g. speech visualisation or speech processing for tactile aids (G10L 15/26 takes precedence)
G10L 2021/065	..	{ Aids for the handicapped in understanding }
G10L 21/10	..	transforming into visible information
G10L 2021/105	...	{ Synthesis of the lips movements from speech, e.g. for talking heads }
G10L 21/12	...	by displaying time domain information
G10L 21/14	...	by displaying frequency domain information
G10L 21/16	..	transforming into a non-visible representation (devices or methods enabling ear patients to replace direct auditory perception by another kind of perception A61F 11/04)
G10L 21/18	..	Details of the transformation process
G10L 25/00		Speech or voice analysis techniques not restricted to a single one of groups G10L 15/00-G10L 21/00
G10L 25/03	.	characterised by the type of extracted parameters
G10L 25/06	..	the extracted parameters being correlation coefficients
G10L 25/09	..	the extracted parameters being zero crossing rates
G10L 25/12	..	the extracted parameters being prediction coefficients
G10L 25/15	..	the extracted parameters being formant information
G10L 25/18	..	the extracted parameters being spectral information of each sub-band
G10L 25/21	..	the extracted parameters being power information
G10L 25/24	..	the extracted parameters being the cepstrum
G10L 25/27	.	characterised by the analysis technique
G10L 25/30	..	using neural networks
G10L 25/33	..	using fuzzy logic
G10L 25/36	..	using chaos theory
G10L 25/39	..	using genetic algorithms
G10L 25/45	.	characterised by the type of analysis window

- G10L 25/48 . specially adapted for particular use
- G10L 25/51 . . for comparison or discrimination
- G10L 25/54 . . . for retrieval
- G10L 25/57 . . . for processing of video signals
- G10L 25/60 . . . for measuring the quality of voice signals
- G10L 25/63 . . . for estimating an emotional state
- G10L 25/66 . . . for extracting parameters related to health condition ([detecting or measuring for diagnostic purposes A61B 5/00](#))
- G10L 25/69 . . for evaluating synthetic or decoded voice signals
- G10L 25/72 . . for transmitting results of analysis
- G10L 25/75 . for modelling vocal tract parameters
- G10L 25/78 . Detection of presence or absence of voice signals ([switching of direction of transmission by voice frequency in two-way loud-speaking telephone systems H04M 9/10](#))
- G10L 2025/783 . . { based on threshold decision }
- G10L 2025/786 . . . { Adaptive threshold }
- G10L 25/81 . . for discriminating voice from music
- G10L 25/84 . . for discriminating voice from noise
- G10L 25/87 . . Detection of discrete points within a voice signal
- G10L 25/90 . Pitch determination of speech signals
- G10L 2025/903 . . { using a laryngograph }
- G10L 2025/906 . . { Pitch tracking }
- G10L 25/93 . Discriminating between voiced and unvoiced parts of speech signals ([G10L 25/90 takes precedence](#))
- G10L 2025/932 . . { Decision in previous or following frames }
- G10L 2025/935 . . { Mixed voiced class; Transitions }
- G10L 2025/937 . . { Signal energy in various frequency bands }
- G10L 99/00** **Subject matter not provided for in other groups of this subclass**