

ECLA EUROPEAN CLASSIFICATION

G10L SPEECH ANALYSIS OR SYNTHESIS; SPEECH RECOGNITION; SPEECH OR VOICE PROCESSING; SPEECH OR AUDIO CODING OR DECODING [C9607] [C1208]

[N: **Notes**

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 [H03M7/30](#).

]

G10L13/00 Speech synthesis; Text to speech systems [N1208]

- G10L13/02 . Methods for producing synthetic speech; Speech synthesisers [N1208]
- G10L13/027 . . Concept to speech synthesisers; Generation of natural phrases from machine-based concepts (generation of parameters for speech synthesis out of text G10L13/08) [N1208]
- G10L13/033 . . Voice editing, e.g. manipulating the voice of the synthesiser [N1208]
- G10L13/033A . . . [N: Pitch control] [N1208]
- G10L13/04 . . Details of speech synthesis systems, e.g. synthesiser structure or memory management [N1208]
- G10L13/04U . . . [N: Synthesisers specially adapted to particular applications] [N9612] [C1208]

[N: **WARNING**

This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L13/00](#) and subgroups. [N1208]

]

- G10L13/047 . . . Architecture of speech synthesisers [N1208]
- G10L13/06 . Elementary speech units used in speech synthesisers; Concatenation rules [N1208]
- G10L13/07 . . Concatenation rules [N1208]
- G10L13/08 . 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 [N1208]
- G10L13/08L . . [N: Detection of language] [N1208]
- G10L13/10 . . Prosody rules derived from text; Stress or intonation [N1208]

G10L15/00 Speech recognition ([G10L17/00](#) takes precedence) [N1208]

- G10L15/00L . [N: Language recognition] [N9612]
- G10L15/01 . Assessment or evaluation of speech recognition systems [N1208]

- G10L15/02 . Feature extraction for speech recognition; Selection of recognition unit [N9607]
- G10L15/04 . Segmentation; Word boundary detection [N1208]
- G10L15/05 . . Word boundary detection [N1208]
- G10L15/06 . Creation of reference templates ; Training of speech recognition systems, e.g. adaptation to the characteristics of the speaker's voice ([G10L15/14 takes precedence](#)) [N1208]
- G10L15/063 . . [N: Training] [N1208]
- G10L15/065 . . Adaptation [N1208]
- G10L15/07 . . . to the speaker [N1208]
- G10L15/07S [N: supervised, i.e. under machine guidance] [N1208]
- G10L15/08 . Speech classification or search [N1208]
- G10L15/08N . . [N: Recognition networks ([G10L15/14M](#), [G10L15/16 take precedence](#))] [N9612]
- G10L15/10 . . using distance or distortion measures between unknown speech and reference templates [N1208]
- G10L15/12 . . using dynamic programming techniques, e.g. dynamic time warping [DTW] [N1208]
- G10L15/14 . . using statistical models, e.g. hidden Markov models [HMMs] ([G10L15/18 takes precedence](#)) [N1208]
- G10L15/14M . . . [N: Hidden Markov Models [HMMs]] [N9607]
- G10L15/14M1 [N: Training of HMMs] [N9607]
- G10L15/14M1S [N: with insufficient amount of training data, e.g. state sharing, tying, deleted interpolation] [N0003]
- G10L15/14M2 [N: Duration modelling in HMMs, e.g. semi HMM, segmental models or transition probabilities [N9607] [C0003]
- G10L15/16 . . using artificial neural networks [N9607] [C0205]
- G10L15/18 . . using natural language modelling [N1208]
- G10L15/18P . . . [N: using prosody or stress] [N1208]
- G10L15/18S . . . [N: Semantic context, e.g. disambiguation of the recognition hypotheses based on word meaning] [N1208]
- G10L15/18U . . . [N: Parsing for meaning understanding] [N1208]
- G10L15/183 . . . using context dependencies, e.g. language models [N1208]
- G10L15/187 Phonemic context, e.g. pronunciation rules, phonotactical constraints or phoneme n-grams [N1208]
- G10L15/19 Grammatical context, e.g. disambiguation of the recognition hypotheses based on word sequence rules [N1208]
- G10L15/193 Formal grammars, e.g. finite state automata, context free grammars or word networks [N1208]
- G10L15/197 Probabilistic grammars, e.g. word n-grams [N1208]
- G10L15/20 . Speech recognition techniques specially adapted for robustness in adverse environments, e.g. in noise, of stress induced speech ([G10L21/02 takes precedence](#)) [N1208]
- G10L15/22 . Procedures used during a speech recognition process, e.g. man-machine dialogue [N1208]

- G10L15/22B . . [N: Barge in, i.e. overridable guidance for interrupting prompts] [N1208]
- G10L15/24 . Speech recognition using non-acoustical features [N1208]
- G10L15/25 . . using position of the lips, movement of the lips or face analysis [N1208]
- G10L15/26 . Speech to text systems ([G10L15/08 takes precedence](#)) [N9607] [C0205]
- G10L15/26A . . [N: Speech recognisers specially adapted for particular applications (devices for signalling identity of wanted subscriber in a telephonic communication equipment controlled by voice recognition [H04M1/27A](#); speech interaction details in interactive information services in a telephonic communication system [H04M3/493S](#))] [N9612][C0503]
- [N: **WARNING**
This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L15/00](#) and subgroups. [N1208]
]
- G10L15/28 . Constructional details of speech recognition systems [N1208]
- G10L15/28H . . [N: Memory allocation or algorithm optimisation to reduce hardware requirements] [N0205]
- G10L15/30 . . Distributed recognition, e.g. in client-server systems, for mobile phones or network applications [N1208]
- G10L15/32 . . Multiple recognisers used in sequence or in parallel; Score combination systems therefor, e.g. voting systems [N1208]
- G10L15/34 . . Adaptation of a single recogniser for parallel processing, e.g. by use of multiple processors or cloud computing [N1208]
- G10L17/00 Speaker identification or verification [N1208]**
- G10L17/00U . . [N: Speaker recognisers specially adapted for particular applications ([G07C9/00B6D takes precedence](#))] [N9612]
- [N: **WARNING**
This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L17/00](#) and subgroups. [N1208]
]
- G10L17/02 . 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 [N1208]
- G10L17/04 . Training, enrolment or model building [N1208]
- G10L17/06 . Decision making techniques; Pattern matching strategies [N1208]
- G10L17/08 . . Use of distortion metrics or a particular distance between probe pattern and reference templates [N1208]
- G10L17/10 . . Multimodal systems, i.e. based on the integration of multiple recognition engines or fusion of expert systems [N1208]
- G10L17/12 . . Score normalisation [N1208]
- G10L17/14 . . Use of phonemic categorisation or speech recognition prior to speaker recognition

- or verification [N1208]
- G10L17/16 . Hidden Markov models [HMMs] [N1208]
- G10L17/18 . Artificial neural networks; Connectionist approaches [N1208]
- G10L17/20 . Pattern transformations or operations aimed at increasing system robustness, e.g. against channel noise or different working conditions [N1208]
- G10L17/22 . Interactive procedures; Man-machine interfaces [N1208]
- G10L17/24 . . the user being prompted to utter a password or a predefined phrase [N1208]
- G10L17/26 . Recognition of special voice characteristics, e.g. for use in lie detectors; Recognition of animal voices [N1208]
- G10L19/00** **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) [N1208]**
- G10L19/00L . [N: Lossless audio signal coding; Perfect reconstruction of coded audio signal by transmission of coding error ([G10L19/24](#) takes precedence)] [N0312] [C1208]
- G10L19/00S . [N: Speech coding using phonetic or linguistic decoding of the source; Reconstruction using text-to-speech synthesis] [N9612] [C0312]
- G10L19/00U . [N: Vocoders specially adapted for particular applications] [N9607]
- [N: **WARNING**
This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L19/00](#) and subgroups. [N1208]
]
- G10L19/002 . Dynamic bit allocation (for perceptual audio coders G10L19/032) [N1208]
- G10L19/005 . Correction of errors induced by the transmission channel, if related to the coding algorithm [N1208]
- G10L19/008 . 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) [N1208]
- G10L19/012 . Comfort noise or silence coding [N1208]
- G10L19/018 . Audio watermarking, i.e. embedding inaudible data in the audio signal [N1208]
- G10L19/02 . using spectral analysis, e.g. transform vocoders or subband vocoders [N1208]
- G10L19/02S . . [N: using subband decomposition] [N9607] [C0312] [C1208]
- G10L19/02S1 . . . [N: Subband vocoders] [N9612]
- G10L19/02T . . [N: using orthogonal transformation] [N1208]
- G10L19/02T2 . . . [N: using wavelet decomposition] [N0312]

- G10L19/022 . . Blocking, i.e. grouping of samples in time; Choice of analysis windows; Overlap factoring [N1208]
- G10L19/025 . . . Detection of transients or attacks for time/frequency resolution switching [N1208]
- G10L19/028 . . Noise substitution, i.e. substituting non-tonal spectral components by noisy source (comfort noise for discontinuous speech transmission G10L19/012) [N1208]
- G10L19/03 . . Spectral prediction for preventing pre-echo; Temporary noise shaping [TNS], e.g. in MPEG2 or MPEG4 [N1208]
- G10L19/032 . . Quantisation or dequantisation of spectral components [N1208]
- G10L19/035 . . . Scalar quantisation [N1208]
- G10L19/038 . . . Vector quantisation, e.g. TwinVQ audio [N1208]

- G10L19/04 . using predictive techniques [N1208]
- G10L19/06 . . Determination or coding of the spectral characteristics, e.g. of the short-term prediction coefficients [N1208]
- G10L19/07 . . . Line spectrum pair [LSP] vocoders [N1208]
- G10L19/08 . . Determination or coding of the excitation function ; Determination or coding of the long-term prediction parameters [N1208]
- G10L19/083 . . . the excitation function being an excitation gain (G10L25/90 takes precedence) [N1208]
- G10L19/087 . . . using mixed excitation models, e.g. MELP, MBE, split band LPC or HVXC [N1208]
- G10L19/09 . . . Long term prediction, i.e. removing periodical redundancies, e.g. by using adaptive codebook or pitch predictor [N1208]
- G10L19/093 . . . using sinusoidal excitation models [N1208]
- G10L19/097 . . . using prototype waveform decomposition or prototype waveform interpolative [PWI] coders [N1208]
- G10L19/10 . . . the excitation function being a multipulse excitation [N1208]
- G10L19/107 Sparse pulse excitation, e.g. by using algebraic codebook [N1208]
- G10L19/113 Regular pulse excitation [N1208]
- G10L19/12 . . . the excitation function being a code excitation, e.g. in code excited linear prediction [CELP] vocoders [N1208]
- G10L19/125 Pitch excitation, e.g. pitch synchronous innovation CELP [PSI-CELP] [N1208]
- G10L19/13 Residual excited linear prediction [RELP] [N1208]
- G10L19/135 Vector sum excited linear prediction [VSELP] [N1208]
- G10L19/16 . . Vocoder architecture [N1208]
- G10L19/167 . . . [N: Audio streaming, i.e. formatting and decoding of an encoded audio signal representation into a data stream for transmission or storage purposes] [N1208]
- G10L19/173 . . . [N: Transcoding, i.e. converting between two coded representations avoiding cascaded coding-decoding] [N1208]
- G10L19/18 . . . Vcoders using multiple modes [N1208]
- G10L19/20 using sound class specific coding, hybrid encoders or object based coding [N1208]
- G10L19/22 Mode decision, i.e. based on audio signal content versus external parameters [N1208]
- G10L19/24 Variable rate codecs, e.g. for generating different qualities using a scalable representation such as hierarchical encoding or layered encoding [N1208]

- G10L19/26 . . Pre-filtering or post-filtering [N1208]
- G10L19/26P . . . [N: Pre-filtering, e.g. high frequency emphasis prior to encoding] [N1208]

- G10L21/00** **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 (G10L19/00 takes precedence) [N1208]**

- G10L21/003 . Changing voice quality, e.g. pitch or formants [N1208]
- G10L21/007 . . characterised by the process used [N1208]
- G10L21/01 . . . Correction of time axis [N1208]
- G10L21/013 . . . Adapting to target pitch [N1208]

- G10L21/02 . Speech enhancement, e.g. noise reduction or echo cancellation (reducing echo effects in line transmission systems [H04B3/20](#) ; echo suppression in hands-free telephones [H04M9/08](#)) [N1208]
- G10L21/02A . . [N: Applications] [N9607]

- [N: **WARNING**
This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L21/00](#) and subgroups. [N1208]
]
- G10L21/02A4 . . . [N: Enhancement of intelligibility of clean or coded speech] [N9607] [C0207]

- [N: **WARNING**
This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to [G10L21/0364](#), [G10L21/057](#). [N1208]
]

- G10L21/0208 . . Noise filtering [N1208]
- G10L21/0216 . . . characterised by the method used for estimating noise [N1208]
- G10L21/0224 Processing in the time domain [N1208]
- G10L21/0232 Processing in the frequency domain [N1208]
- G10L21/0264 . . . characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques [N1208]
- G10L21/0272 . . Voice signal separating [N1208]
- G10L21/028 . . . using properties of sound source [N1208]
- G10L21/0308 . . . characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques [N1208]
- G10L21/0316 . . by changing the amplitude [N1208]
- G10L21/0324 . . . Details of processing therefor [N1208]
- G10L21/0332 involving modification of waveforms [N1208]
- G10L21/034 Automatic adjustment [N1208]
- G10L21/0356 . . . for synchronising with other signals, e.g. video signals [N1208]
- G10L21/0364 . . . for improving intelligibility [N1208]
- G10L21/038 . . using band spreading techniques [N1208]
- G10L21/0388 . . . Details of processing therefor [N1208]

- G10L21/04 . Time compression or expansion [N1208]

- G10L21/043 . . by changing speed [N1208]
- G10L21/045 . . . using thinning out or insertion of a waveform [N1208]
- G10L21/047 characterised by the type of waveform to be thinned out or inserted [N1208]
- G10L21/049 characterised by the interconnection of waveforms [N1208]
- G10L21/055 . . for synchronising with other signals, e.g. video signals [N1208]
- G10L21/057 . . for improving intelligibility [N1208]

- G10L21/06 . Transformation of speech into a non-audible representation, e.g. speech visualisation or speech processing for tactile aids ([G10L15/26](#) takes precedence) [N1208]
- G10L21/10 . . transforming into visible information [N1208]
- G10L21/12 . . . by displaying time domain information [N1208]
- G10L21/14 . . . by displaying frequency domain information [N1208]
- G10L21/16 . . 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) [N1208]
- G10L21/18 . . Details of the transformation process [N1208]

G10L25/00 **Speech or voice analysis techniques not restricted to a single one of groups G10L15/00-G10L21/00** [N1208]

- G10L25/03 . characterised by the type of extracted parameters [N1208]
- G10L25/06 . . the extracted parameters being correlation coefficients [N1208]
- G10L25/09 . . the extracted parameters being zero crossing rates [N1208]
- G10L25/12 . . the extracted parameters being prediction coefficients [N1208]
- G10L25/15 . . the extracted parameters being formant information [N1208]
- G10L25/18 . . the extracted parameters being spectral information of each sub-band [N1208]
- G10L25/21 . . the extracted parameters being power information [N1208]
- G10L25/24 . . the extracted parameters being the cepstrum [N1208]

- G10L25/27 . characterised by the analysis technique [N1208]
- G10L25/30 . . using neural networks [N1208]
- G10L25/33 . . using fuzzy logic [N1208]
- G10L25/36 . . using chaos theory [N1208]
- G10L25/39 . . using genetic algorithms [N1208]

- G10L25/45 . characterised by the type of analysis window [N1208]

- G10L25/48 . specially adapted for particular use [N1208]
- G10L25/51 . . for comparison or discrimination [N1208]
- G10L25/54 . . . for retrieval [N1208]
- G10L25/57 . . . for processing of video signals [N1208]
- G10L25/60 . . . for measuring the quality of voice signals [N1208]
- G10L25/63 . . . for estimating an emotional state [N1208]
- G10L25/66 . . . for extracting parameters related to health condition ([detecting or measuring for diagnostic purposes A61B5/00](#)) [N1208]

- G10L25/69 . . for evaluating synthetic or decoded voice signals [N1208]
- G10L25/72 . . for transmitting results of analysis [N1208]
- G10L25/75 . for modelling vocal tract parameters [N1208]
- G10L25/78 . Detection of presence or absence of voice signals (switching of direction of transmission by voice frequency in two-way loud-speaking telephone systems [H04M9/10](#)) [N1208]
- G10L25/81 . . for discriminating voice from music [N1208]
- G10L25/84 . . for discriminating voice from noise [N1208]
- G10L25/87 . . Detection of discrete points within a voice signal [N1208]
- G10L25/90 . Pitch determination of speech signals [N1208]
- G10L25/93 . Discriminating between voiced and unvoiced parts of speech signals ([G10L25/90 takes precedence](#)) [N1208]

- G10L99/00** **Subject matter not provided for in other groups of this subclass** [N1208]