

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](#).

| | | | |
|------------------|---|------------------|---|
| 13/00 | Speech synthesis; Text to speech systems | 2015/0635 | . . . {updating or merging of old and new templates; Mean values; Weighting} |
| 13/02 | . Methods for producing synthetic speech; Speech synthesisers | 2015/0636 | {Threshold criteria for the updating} |
| 2013/021 | . . {Overlap-add techniques} | 2015/0638 | . . . {Interactive procedures} |
| 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) | 15/065 | . . Adaptation |
| 13/033 | . . Voice editing, e.g. manipulating the voice of the synthesiser | 15/07 | . . . to the speaker |
| 13/0335 | . . . {Pitch control} | 15/075 | {supervised, i.e. under machine guidance} |
| 13/04 | . . Details of speech synthesis systems, e.g. synthesiser structure or memory management | 15/08 | . Speech classification or search |
| 13/043 | . . . {Synthesisers specially adapted to particular applications} | 2015/081 | . . {Search algorithms, e.g. Baum-Welch or Viterbi} |
| 13/047 | . . . Architecture of speech synthesisers | 15/083 | . . {Recognition networks (G10L 15/142 , G10L 15/16 take precedence)} |
| 13/06 | . Elementary speech units used in speech synthesisers; Concatenation rules | 2015/085 | . . {Methods for reducing search complexity, pruning} |
| 13/07 | . . Concatenation rules | 2015/086 | . . {Recognition of spelled words} |
| 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 | 2015/088 | . . {Word spotting} |
| 2013/083 | . . {Special characters, e.g. punctuation marks} | 15/10 | . . using distance or distortion measures between unknown speech and reference templates |
| 13/086 | . . {Detection of language} | 15/12 | . . using dynamic programming techniques, e.g. dynamic time warping [DTW] |
| 13/10 | . . Prosody rules derived from text; Stress or intonation | 15/14 | . . using statistical models, e.g. hidden Markov models [HMMs] (G10L 15/18 takes precedence) |
| 2013/105 | . . . {Duration} | 15/142 | . . . {Hidden Markov Models [HMMs]} |
| 15/00 | Speech recognition (G10L 17/00 takes precedence) | 15/144 | {Training of HMMs} |
| 15/005 | . {Language recognition} | 15/146 | {with insufficient amount of training data, e.g. state sharing, tying, deleted interpolation} |
| 15/01 | . Assessment or evaluation of speech recognition systems | 15/148 | {Duration modelling in HMMs, e.g. semi HMM, segmental models or transition probabilities} |
| 15/02 | . Feature extraction for speech recognition; Selection of recognition unit | 15/16 | . . using artificial neural networks |
| 2015/022 | . . {Demisyllables, biphones or triphones being the recognition units} | 15/18 | . . using natural language modelling |
| 2015/025 | . . {Phonemes, fenemes or fenones being the recognition units} | 15/1807 | . . . {using prosody or stress} |
| 2015/027 | . . {Syllables being the recognition units} | 15/1815 | . . . {Semantic context, e.g. disambiguation of the recognition hypotheses based on word meaning} |
| 15/04 | . Segmentation; Word boundary detection | 15/1822 | . . . {Parsing for meaning understanding} |
| 15/05 | . . Word boundary detection | 15/183 | . . . using context dependencies, e.g. language models |
| 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) | 15/187 | Phonemic context, e.g. pronunciation rules, phonotactical constraints or phoneme n-grams |
| 15/063 | . . {Training} | 15/19 | Grammatical context, e.g. disambiguation of the recognition hypotheses based on word sequence rules |
| 2015/0631 | . . . {Creating reference templates; Clustering} | 15/193 | Formal grammars, e.g. finite state automata, context free grammars or word networks |
| 2015/0633 | {using lexical or orthographic knowledge sources} | 15/197 | Probabilistic grammars, e.g. word n-grams |

| | | | |
|--------------|---|--------------|---|
| 15/20 | <ul style="list-style-type: none"> Speech recognition techniques specially adapted for robustness in adverse environments, e.g. in noise, of stress induced speech (G10L 21/02 takes precedence) | 17/20 | <ul style="list-style-type: none"> Pattern transformations or operations aimed at increasing system robustness, e.g. against channel noise or different working conditions |
| 15/22 | <ul style="list-style-type: none"> Procedures used during a speech recognition process, e.g. man-machine dialogue | 17/22 | <ul style="list-style-type: none"> Interactive procedures; Man-machine interfaces |
| 2015/221 | <ul style="list-style-type: none"> . . {Announcement of recognition results} | 17/24 | <ul style="list-style-type: none"> . . the user being prompted to utter a password or a predefined phrase |
| 15/222 | <ul style="list-style-type: none"> . . {Barge in, i.e. overridable guidance for interrupting prompts} | 17/26 | <ul style="list-style-type: none"> Recognition of special voice characteristics, e.g. for use in lie detectors; Recognition of animal voices |
| 2015/223 | <ul style="list-style-type: none"> . . {Execution procedure of a spoken command} | 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) |
| 2015/225 | <ul style="list-style-type: none"> . . {Feedback of the input speech} | 2019/0001 | <ul style="list-style-type: none"> . {Codebooks} |
| 2015/226 | <ul style="list-style-type: none"> . . {Taking into account non-speech characteristics} | 2019/0002 | <ul style="list-style-type: none"> . . {Codebook adaptations} |
| 2015/227 | <ul style="list-style-type: none"> . . . {of the speaker; Human-factor methodology} | 2019/0003 | <ul style="list-style-type: none"> . . {Backward prediction of gain} |
| 2015/228 | <ul style="list-style-type: none"> . . . {of application context} | 2019/0004 | <ul style="list-style-type: none"> . . {Design or structure of the codebook} |
| 15/24 | <ul style="list-style-type: none"> Speech recognition using non-acoustical features | 2019/0005 | <ul style="list-style-type: none"> . . . {Multi-stage vector quantisation} |
| 15/25 | <ul style="list-style-type: none"> . . using position of the lips, movement of the lips or face analysis | 2019/0006 | <ul style="list-style-type: none"> . . . {Tree or treillis structures; Delayed decisions} |
| 15/26 | <ul style="list-style-type: none"> Speech to text systems (G10L 15/08 takes precedence) | 2019/0007 | <ul style="list-style-type: none"> . . {Codebook element generation} |
| 15/265 | <ul style="list-style-type: none"> . . {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)} | 2019/0008 | <ul style="list-style-type: none"> . . . {Algebraic codebooks} |
| 15/28 | <ul style="list-style-type: none"> Constructional details of speech recognition systems | 2019/0009 | <ul style="list-style-type: none"> . . . {Orthogonal codebooks} |
| 15/285 | <ul style="list-style-type: none"> . . {Memory allocation or algorithm optimisation to reduce hardware requirements} | 2019/001 | <ul style="list-style-type: none"> . . . {Interpolation of codebook vectors} |
| 15/30 | <ul style="list-style-type: none"> . . Distributed recognition, e.g. in client-server systems, for mobile phones or network applications | 2019/0011 | <ul style="list-style-type: none"> . . {Long term prediction filters, i.e. pitch estimation} |
| 15/32 | <ul style="list-style-type: none"> . . Multiple recognisers used in sequence or in parallel; Score combination systems therefor, e.g. voting systems | 2019/0012 | <ul style="list-style-type: none"> . . {Smoothing of parameters of the decoder interpolation} |
| 15/34 | <ul style="list-style-type: none"> . . Adaptation of a single recogniser for parallel processing, e.g. by use of multiple processors or cloud computing | 2019/0013 | <ul style="list-style-type: none"> . . {Codebook search algorithms} |
| 17/00 | Speaker identification or verification | 2019/0014 | <ul style="list-style-type: none"> . . . {Selection criteria for distances} |
| 17/005 | <ul style="list-style-type: none"> . {Speaker recognisers specially adapted for particular applications (G07C 9/00071 takes precedence)} | 2019/0015 | <ul style="list-style-type: none"> . . . {Viterbi algorithms} |
| 17/02 | <ul style="list-style-type: none"> 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 | 2019/0016 | <ul style="list-style-type: none"> . . {Codebook for LPC parameters} |
| 17/04 | <ul style="list-style-type: none"> Training, enrolment or model building | 19/0017 | <ul style="list-style-type: none"> . {Lossless audio signal coding; Perfect reconstruction of coded audio signal by transmission of coding error (G10L 19/24 takes precedence)} |
| 17/06 | <ul style="list-style-type: none"> Decision making techniques; Pattern matching strategies | 19/0018 | <ul style="list-style-type: none"> . {Speech coding using phonetic or linguistic decoding of the source; Reconstruction using text-to-speech synthesis} |
| 17/08 | <ul style="list-style-type: none"> . . Use of distortion metrics or a particular distance between probe pattern and reference templates | 19/0019 | <ul style="list-style-type: none"> . {Vocoders specially adapted for particular applications} |
| 17/10 | <ul style="list-style-type: none"> . . Multimodal systems, i.e. based on the integration of multiple recognition engines or fusion of expert systems | 19/002 | <ul style="list-style-type: none"> . Dynamic bit allocation (for perceptual audio coders G10L 19/032) |
| 17/12 | <ul style="list-style-type: none"> . . Score normalisation | 19/005 | <ul style="list-style-type: none"> . Correction of errors induced by the transmission channel, if related to the coding algorithm |
| 17/14 | <ul style="list-style-type: none"> . . Use of phonemic categorisation or speech recognition prior to speaker recognition or verification | 19/008 | <ul style="list-style-type: none"> . 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) |
| 17/16 | <ul style="list-style-type: none"> Hidden Markov models [HMMs] | 19/012 | <ul style="list-style-type: none"> . Comfort noise or silence coding |
| 17/18 | <ul style="list-style-type: none"> Artificial neural networks; Connectionist approaches | 19/018 | <ul style="list-style-type: none"> . Audio watermarking, i.e. embedding inaudible data in the audio signal |
| | | 19/02 | <ul style="list-style-type: none"> . using spectral analysis, e.g. transform vocoders or subband vocoders |
| | | 19/0204 | <ul style="list-style-type: none"> . . {using subband decomposition} |
| | | 19/0208 | <ul style="list-style-type: none"> . . . {Subband vocoders} |
| | | 19/0212 | <ul style="list-style-type: none"> . . {using orthogonal transformation} |
| | | 19/0216 | <ul style="list-style-type: none"> . . . {using wavelet decomposition} |
| | | 19/022 | <ul style="list-style-type: none"> . . Blocking, i.e. grouping of samples in time; Choice of analysis windows; Overlap factoring |

| | | | |
|--------|---|------------|--|
| 19/025 | . . . Detection of transients or attacks for time/frequency resolution switching | 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) |
| 19/028 | . . Noise substitution, i.e. substituting non-tonal spectral components by noisy source (comfort noise for discontinuous speech transmission G10L 19/012) | 21/003 | . Changing voice quality, e.g. pitch or formants |
| 19/03 | . . Spectral prediction for preventing pre-echo; Temporary noise shaping [TNS], e.g. in MPEG2 or MPEG4 | 21/007 | . . characterised by the process used |
| 19/032 | . . Quantisation or dequantisation of spectral components | 21/01 | . . . Correction of time axis |
| 19/035 | . . . Scalar quantisation | 21/013 | . . . Adapting to target pitch |
| 19/038 | . . . Vector quantisation, e.g. TwinVQ audio | 2021/0135 | {Voice conversion or morphing} |
| 19/04 | . using predictive techniques | 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) |
| 19/06 | . . Determination or coding of the spectral characteristics, e.g. of the short-term prediction coefficients | 21/0202 | . . {Applications} |
| 19/07 | . . . Line spectrum pair [LSP] vocoders | 21/0205 | . . . {Enhancement of intelligibility of clean or coded speech} |
| 19/08 | . . Determination or coding of the excitation function; Determination or coding of the long-term prediction parameters | 21/0208 | . . Noise filtering |
| 19/083 | . . . the excitation function being an excitation gain (G10L 25/90 takes precedence) | 2021/02082 | . . . {the noise being echo, reverberation of the speech} |
| 19/087 | . . . using mixed excitation models, e.g. MELP, MBE, split band LPC or HVXC | 2021/02085 | . . . {Periodic noise} |
| 19/09 | . . . Long term prediction, i.e. removing periodical redundancies, e.g. by using adaptive codebook or pitch predictor | 2021/02087 | . . . {the noise being separate speech, e.g. cocktail party} |
| 19/093 | . . . using sinusoidal excitation models | 21/0216 | . . . characterised by the method used for estimating noise |
| 19/097 | . . . using prototype waveform decomposition or prototype waveform interpolative [PWI] coders | 2021/02161 | {Number of inputs available containing the signal or the noise to be suppressed} |
| 19/10 | . . . the excitation function being a multipulse excitation | 2021/02163 | {Only one microphone} |
| 19/107 | Sparse pulse excitation, e.g. by using algebraic codebook | 2021/02165 | {Two microphones, one receiving mainly the noise signal and the other one mainly the speech signal} |
| 19/113 | Regular pulse excitation | 2021/02166 | {Microphone arrays; Beamforming} |
| 19/12 | . . . the excitation function being a code excitation, e.g. in code excited linear prediction [CELP] vocoders | 2021/02168 | {the estimation exclusively taking place during speech pauses} |
| 19/125 | Pitch excitation, e.g. pitch synchronous innovation CELP [PSI-CELP] | 21/0224 | Processing in the time domain |
| 19/13 | Residual excited linear prediction [RELP] | 21/0232 | Processing in the frequency domain |
| 19/135 | Vector sum excited linear prediction [VSELP] | 21/0264 | . . . characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques |
| 19/16 | . . Vocoder architecture | 21/0272 | . . Voice signal separating |
| 19/167 | . . . {Audio streaming, i.e. formatting and decoding of an encoded audio signal representation into a data stream for transmission or storage purposes} | 21/028 | . . . using properties of sound source |
| 19/173 | . . . {Transcoding, i.e. converting between two coded representations avoiding cascaded coding-decoding} | 21/0308 | . . . characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques |
| 19/18 | . . . Vocoders using multiple modes | 21/0316 | . . by changing the amplitude |
| 19/20 | using sound class specific coding, hybrid encoders or object based coding | 21/0324 | . . . Details of processing therefor |
| 19/22 | Mode decision, i.e. based on audio signal content versus external parameters | 21/0332 | involving modification of waveforms |
| 19/24 | Variable rate codecs, e.g. for generating different qualities using a scalable representation such as hierarchical encoding or layered encoding | 21/034 | Automatic adjustment |
| 19/26 | . . Pre-filtering or post-filtering | 21/0356 | . . . for synchronising with other signals, e.g. video signals |
| 19/265 | . . . {Pre-filtering, e.g. high frequency emphasis prior to encoding} | 21/0364 | . . . for improving intelligibility |
| | | 2021/03643 | {Diver speech} |
| | | 2021/03646 | {Stress or Lombard effect} |
| | | 21/038 | . . using band spreading techniques |
| | | 21/0388 | . . . Details of processing therefor |
| | | 21/04 | . Time compression or expansion |
| | | 21/043 | . . by changing speed |
| | | 21/045 | . . . using thinning out or insertion of a waveform |
| | | 21/047 | characterised by the type of waveform to be thinned out or inserted |
| | | 21/049 | characterised by the interconnection of waveforms |
| | | 21/055 | . . for synchronising with other signals, e.g. video signals |

| | | | |
|--------------|--|----------|--|
| 21/057 | . . for improving intelligibility | 25/93 | . Discriminating between voiced and unvoiced parts of speech signals (G10L 25/90 takes precedence) |
| 2021/0575 | . . . {Aids for the handicapped in speaking} | 2025/932 | . . {Decision in previous or following frames} |
| 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) | 2025/935 | . . {Mixed voiced class; Transitions} |
| 2021/065 | . . . {Aids for the handicapped in understanding} | 2025/937 | . . {Signal energy in various frequency bands} |
| 21/10 | . . transforming into visible information | 99/00 | Subject matter not provided for in other groups of this subclass |
| 2021/105 | . . . {Synthesis of the lips movements from speech, e.g. for talking heads} | | |
| 21/12 | . . . by displaying time domain information | | |
| 21/14 | . . . by displaying frequency domain information | | |
| 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) | | |
| 21/18 | . . Details of the transformation process | | |
| 25/00 | Speech or voice analysis techniques not restricted to a single one of groups G10L 15/00-G10L 21/00 | | |
| 25/03 | . characterised by the type of extracted parameters | | |
| 25/06 | . . the extracted parameters being correlation coefficients | | |
| 25/09 | . . the extracted parameters being zero crossing rates | | |
| 25/12 | . . the extracted parameters being prediction coefficients | | |
| 25/15 | . . the extracted parameters being formant information | | |
| 25/18 | . . the extracted parameters being spectral information of each sub-band | | |
| 25/21 | . . the extracted parameters being power information | | |
| 25/24 | . . the extracted parameters being the cepstrum | | |
| 25/27 | . characterised by the analysis technique | | |
| 25/30 | . . using neural networks | | |
| 25/33 | . . using fuzzy logic | | |
| 25/36 | . . using chaos theory | | |
| 25/39 | . . using genetic algorithms | | |
| 25/45 | . characterised by the type of analysis window | | |
| 25/48 | . specially adapted for particular use | | |
| 25/51 | . . for comparison or discrimination | | |
| 25/54 | . . . for retrieval | | |
| 25/57 | . . . for processing of video signals | | |
| 25/60 | . . . for measuring the quality of voice signals | | |
| 25/63 | . . . for estimating an emotional state | | |
| 25/66 | . . . for extracting parameters related to health condition (detecting or measuring for diagnostic purposes A61B 5/00) | | |
| 25/69 | . . for evaluating synthetic or decoded voice signals | | |
| 25/72 | . . for transmitting results of analysis | | |
| 25/75 | . for modelling vocal tract parameters | | |
| 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) | | |
| 2025/783 | . . {based on threshold decision} | | |
| 2025/786 | . . . {Adaptive threshold} | | |
| 25/81 | . . for discriminating voice from music | | |
| 25/84 | . . for discriminating voice from noise | | |
| 25/87 | . . Detection of discrete points within a voice signal | | |
| 25/90 | . Pitch determination of speech signals | | |
| 2025/903 | . . {using a laryngograph} | | |
| 2025/906 | . . {Pitch tracking} | | |