## CPC - COOPERATIVE PATENT CLASSIFICATION

**G** PHYSICS  
*(NOTES omitted)*

### INSTRUMENTS

**G10** MUSICAL INSTRUMENTS; ACOUSTICS  
*(NOTES omitted)*

**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 or audio 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.

**WARNING**
In this subclass non-limiting references (in the sense of paragraph 39 of the Guide to the IPC) may still be displayed in the scheme.

### 13/00 Speech synthesis; Text to speech systems
### 13/02 . Methods for producing synthetic speech; Speech synthesisers

- 2013/021 . {Overlap-add techniques}

### 13/03 Voice editing, e.g. manipulating the voice of the synthesiser
### 13/033 .

### 13/035 . {Pitch control}
### 13/04 Details of speech synthesis systems, e.g. synthesiser structure or memory management
### 13/047 . Architecture of speech synthesisers
### 13/06 . Elementary speech units used in speech synthesisers; Concatenation rules
### 13/07 . Concatenation rules
### 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

- 2013/083 . {Special characters, e.g. punctuation marks}
### 13/086 . {Detection of language}
### 13/10 . Prosody rules derived from text; Stress or intonation

- 2013/105 . {Duration}

### 15/00 Speech recognition *(G10L 17/00 takes precedence)*
### 15/005 . {Language recognition}
### 15/01 . Assessment or evaluation of speech recognition systems
### 15/02 . Feature extraction for speech recognition; Selection of recognition unit

- 2015/022 . {Demisyllables, biphones or triphones being the recognition units}
### 15/025 . {Phonemes, fenesmes or fenomes being the recognition units}
### 15/027 . {Syllables being the recognition units}

### 15/04 . Segmentation; Word boundary detection
### 15/05 . Word boundary detection
### 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)*
### 15/063 . {Training}
### 2015/0631 . {Creating reference templates; Clustering}
### 2015/0633 . {using lexical or orthographic knowledge sources}
### 2015/0635 . {updating or merging of old and new templates; Mean values; Weighting}
### 2015/0636 . {Threshold criteria for the updating}
### 2015/0638 . {Interactive procedures}
### 15/065 . Adaptation
### 15/07 . to the speaker
### 15/075 . {supervised, i.e. under machine guidance}
### 15/08 . Speech classification or search
### 2015/081 . {Search algorithms, e.g. Baum-Welch or Viterbi}
### 15/083 . {Recognition networks *(G10L 15/142, G10L 15/16 take precedence)*}
### 2015/085 . {Methods for reducing search complexity, pruning}
### 2015/086 . {Recognition of spelled words}
### 2015/088 . {Word spotting}
### 15/10 . using distance or distortion measures between unknown speech and reference templates
### 15/12 . using dynamic programming techniques, e.g. dynamic time warping [DTW]
### 15/14 . using statistical models, e.g. Hidden Markov Models [HMMs] *(G10L 15/18 takes precedence)*
### 15/142 . {Hidden Markov Models [HMMs]}
### 15/144 . {Training of HMMs}
### 15/146 . . . . {with insufficient amount of training data, e.g. state sharing, tying, deleted interpolation}
17/00 Speaker identification or verification
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
17/04 Training, enrolment or model building
17/06 Decision making techniques; Pattern matching strategies
17/08 Use of distortion metrics or a particular distance between probe pattern and reference templates

19/00 Speech or audio signals analysis-synthesis techniques for redundancies reduction, e.g. in vocoders; Coding or decoding of speech or audio signals, using source filter models or psychoacoustic analysis (in musical instruments G10H)

19/001 [Codebooks]
19/002 [Codebook adaptations]
19/003 [Backward prediction of gain]
19/004 [Design or structure of the codebook]
19/005 [Multi-stage vector quantisation]
19/006 [Tree or trellis structures; Delayed decisions]
19/007 [Codebook element generation]
19/008 [Algebraic codebooks]
19/009 [Orthogonal codebooks]
19/01 [Interpolation of codebook vectors]
19/011 [Long term prediction filters, i.e. pitch estimation]
19/012 [Smoothing of parameters of the decoder interpolation]
19/013 [Codebook search algorithms]
19/014 [Selection criteria for distances]
19/015 [Viterbi algorithms]
19/016 [Codebook for LPC parameters]
19/017 [Lossless audio signal coding; Perfect reconstruction of coded audio signal by transmission of coding error (G10L 19/24 takes precedence)]
19/018 [Speech coding using phonetic or linguistic decoding of the source; Reconstruction using text-to-speech synthesis]
19/02 Dynamic bit allocation (for perceptual audio coders G10L 19/032)
19/05 Correction of errors induced by the transmission channel, if related to the coding algorithm
19/08 Multichannel audio signal coding or decoding using interchannel correlation to reduce redundancy, e.g. joint-stereo, intensity-coding or matrixing
19/012 Comfort noise or silence coding
19/018 Audio watermarking, i.e. embedding inaudible data in the audio signal
19/02 Using spectral analysis, e.g. transform vocoders or subband vocoders
19/0204 [using subband decomposition]
19/0208 [Subband vocoders]
19/0212 . . . [using orthogonal transformation]
19/0216 . . . [using wavelet decomposition]
19/0222 . . . Blocking, i.e. grouping of samples in time; Choice of analysis windows; Overlap factoring
19/0225 . . . Detection of transients or attacks for time/ frequency resolution switching
19/0228 . . . Noise substitution, i.e. substituting non-tonal spectral components by noisy source (comfort noise for discontinuous speech transmission G10L 19/012)
19/03 . . . Spectral prediction for preventing pre-echo; Temporary noise shaping [TNS], e.g. in MPEG2 or MPEG4
19/032 . . . Quantisation or dequantisation of spectral components
19/035 . . . Scalar quantisation
19/038 . . . Vector quantisation, e.g. TwinVQ audio
19/04 . . . using predictive techniques
19/06 . . . Determination or coding of the spectral characteristics, e.g. of the short-term prediction coefficients
19/07 . . . Line spectrum pair [LSP] vocoders
19/08 . . . Determination or coding of the excitation function; Determination or coding of the long-term prediction parameters
19/083 . . . the excitation function being an excitation gain (G10L 25/90 takes precedence)
19/087 . . . using mixed excitation models, e.g. MELP, MBE, split band LPC or HVXC
19/09 . . . Long term prediction, i.e. removing periodical redundancies, e.g. by using adaptive codebook or pitch predictor
19/093 . . . using sinusoidal excitation models
19/097 . . . using prototype waveform decomposition or prototype waveform interpolative [PWI] coders
19/10 . . . the excitation function being a multipulse excitation
19/107 . . . Sparse pulse excitation, e.g. by using algebraic codebook
19/113 . . . Regular pulse excitation
19/12 . . . the excitation function being a code excitation, e.g. in code excited linear prediction [CELP] vocoders
19/125 . . . Pitch excitation, e.g. pitch synchronous innovation CELP [PSI-CELP]
19/13 . . . Residual excited linear prediction [RELP]
19/135 . . . Vector sum excited linear prediction [VSELP]
19/16 . . . Vocoder architecture
19/167 . . . [Audio streaming, i.e. formatting and decoding of an encoded audio signal representation into a data stream for transmission or storage purposes]
19/173 . . . [Transcoding, i.e. converting between two coded representations avoiding cascaded coding-decoding]
19/18 . . . Vocoder using multiple modes
19/20 . . . using sound class specific coding, hybrid encoders or object based coding
19/22 . . . Mode decision, i.e. based on audio signal content versus external parameters
19/24 . . . . . . Variable rate codecs, e.g. for generating different qualities using a scalable representation such as hierarchical encoding or layered encoding
19/26 . . . Pre-filtering or post-filtering
19/265 . . . [Pre-filtering, e.g. high frequency emphasis prior to encoding]

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)
21/003 . . . Changing voice quality, e.g. pitch or formants
21/007 . . . characterised by the process used
21/01 . . . Correction of time axis
21/013 . . . Adapting to target pitch
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)

WARNING
Group G10L 21/02 is incomplete pending reclassification of documents from group G10L 21/0202.

Groups G10L 21/0202 and G10L 21/02 should be considered in order to perform a complete search.

21/0202 (Frozen) . . . [Applications]

WARNING
Group G10L 21/0202 is no longer used for the classification of documents as of August 1, 2020.
The content of this group is being reclassified into groups G10L 21/02, G10L 21/0316, G10L 21/0354, G10L 2021/03643, and G10L 2021/03646.

All groups listed in this Warning should be considered in order to perform a complete search.

21/0208 . . . Noise filtering
21/02082 . . . [the noise being echo, reverberation of the speech]
21/02085 . . . [Periodic noise]
21/02087 . . . [the noise being separate speech, e.g. cocktail party]
21/0216 . . . characterised by the method used for estimating noise
21/02161 . . . [Number of inputs available containing the signal or the noise to be suppressed]
21/02163 . . . . . . [Only one microphone]
21/02165 . . . . . . [Two microphones, one receiving mainly the noise signal and the other one mainly the speech signal]
21/02166 . . . . . . [Microphone arrays; Beamforming]
21/02168 . . . . . . [the estimation exclusively taking place during speech pauses]
21/0224 . . . Processing in the time domain
21/0232 . . . Processing in the frequency domain

CPC - 2021.01
Time compression or expansion for improving intelligibility signals using band spreading techniques and G10L 21/0316

WARNING

Group G10L 21/0316 is incomplete pending reclassification of documents from group G10L 21/0202.

Groups G10L 21/0202 and G10L 21/0316 should be considered in order to perform a complete search.

Details of processing therefor

21/0324

21/034

for automatic adjustment

21/0356

for improving synchronising with other signals, e.g. video signals

21/0364

for improving intelligibility

WARNING

Group G10L 21/0364 is incomplete pending reclassification of documents from group G10L 21/0202.

Groups G10L 21/0202 and G10L 21/0364 should be considered in order to perform a complete search.

25/00 Speech or voice analysis techniques not restricted to a single one of groups G10L 15/00 - G10L 21/00 (muting semiconductor-based amplifiers when some special characteristics of a signal are sensed by a speech detector, e.g. sensing when no signal is present, H03G 3/34)

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)

25/783 [based on threshold decision]

25/786 [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

25/903 [using a laryngograph]
G10L

2025/906 . . {Pitch tracking}
25/93 . Discriminating between voiced and unvoiced parts of speech signals (G10L 25/90 takes precedence)
2025/932 . . {Decision in previous or following frames}
2025/935 . . {Mixed voiced class; Transitions}
2025/937 . . {Signal energy in various frequency bands}

99/00 Subject matter not provided for in other groups of this subclass