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}
13/027 . Concept to speech synthesisers; Generation of natural phrases from machine-based concepts (generation of parameters for speech synthesis out of text G10L 13/08)
13/033 . Voice editing, e.g. manipulating the voice of the synthesiser
13/0335 . {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 . {Demi-syllables, biphones or triphones being the recognition units}
2015/025 . {Phonemes, fenemes or fenones being the recognition units}
2015/027 . {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}
Speaker identification or verification

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

Training, enrolment or model building

Decision making techniques; Pattern matching strategies

Use of distortion metrics or a particular distance between probe pattern and reference templates

17/10  .  .  .  .  .  Multimodal systems, i.e. based on the integration of multiple recognition engines or fusion of expert systems

17/12  .  .  .  .  .  Score normalisation

17/14  .  .  .  .  .  Use of phonemic categorisation or speech recognition prior to speaker recognition or verification

17/16  .  .  .  .  .  Hidden Markov models [HMMs]

17/18  .  .  .  .  .  Artificial neural networks; Connectionist approaches

17/20  .  .  .  .  .  Pattern transformations or operations aimed at increasing system robustness, e.g. against channel noise or different working conditions

17/22  .  .  .  .  .  Interactive procedures; Man-machine interfaces

17/24  .  .  .  .  .  the user being prompted to utter a password or a predefined phrase

17/26  .  .  .  .  .  Recognition of special voice characteristics, e.g. for use in lie detectors; Recognition of animal voices

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

2019/0001  .  [Codebooks]

2019/0002  .  .  [Codebook adaptations]

2019/0003  .  .  [Backward prediction of gain]

2019/0004  .  .  [Design or structure of the codebook]

2019/0005  .  .  [Multi-stage vector quantisation]

2019/0006  .  .  [Tree or trellis structures; Delayed decisions]

2019/0007  .  .  [Codebook element generation]

2019/0008  .  .  [Algebraic codebooks]

2019/0009  .  .  [Orthogonal codebooks]

2019/001  .  .  .  [Interpolation of codebook vectors]

2019/0011  .  .  .  [Long term prediction filters, i.e. pitch estimation]

2019/0012  .  .  .  [Smoothing of parameters of the decoder interpolation]

2019/0013  .  .  [Codebook search algorithms]

2019/0014  .  .  .  [Selection criteria for distances]

2019/0015  .  .  [Viterbi algorithms]

2019/0016  .  .  [Codebook for LPC parameters]

19/0017  .  .  .  .  .  Lossless audio signal coding; Perfect reconstruction of coded audio signal by transmission of coding error (G10L 19/24 takes precedence)

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

19/002  .  .  .  .  .  Dynamic bit allocation (for perceptual audio coders G10L 19/032)

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

19/008  .  .  .  .  .  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]
Using predictive techniques, determination or coding of the long-term prediction parameters, quantisation or dequantisation of spectral components, scalar quantisation, vector quantisation, e.g. twin VQ audio, determination or coding of the spectral characteristics, e.g. of the short-term prediction coefficients, line spectrum pair [LSP] vocoders, determination or coding of the excitation function; determination or coding of the long-term prediction parameters, the excitation function being an excitation gain, e.g. in MPEG2 or MPEG4, using mixed excitation models, e.g. MELP, MBE, split band LPC or HVXC, long term prediction, i.e. removing periodical redundancies, e.g. by using adaptive codebook or pitch predictor, using sinusoidal excitation models, using prototype waveform decomposition or prototype waveform interpolative [PWI] coders, the excitation function being a multipulse excitation, sparse pulse excitation, e.g. by using algebraic codebook, regular pulse excitation, the excitation function being a code excitation, e.g. in code excited linear prediction [CELP] vocoders, pitch excitation, e.g. pitch synchronous innovation CELP [PSI-CELP], residual excited linear prediction [REL P], vector sum excited linear prediction [VSELP], vocoder architecture, audio streaming, i.e. formatting and decoding of an encoded audio signal representation into a data stream for transmission or storage purposes, transcoding, i.e. converting between two coded representations avoiding cascaded coding-decoding, vocoders using multiple modes, using sound class specific coding, hybrid encoders or object based coding, mode decision, i.e. based on audio signal content versus external parameters, variable rate codecs, e.g. for generating different qualities using a scalable representation such as hierarchical encoding or layered encoding, pre-filtering or post-filtering, [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)

- Changing voice quality, e.g. pitch or formants
- characterised by the process used
- Correction of time axis
- Adapting to target pitch

**2021/0135**

- 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.

**2021/0202** *(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 21/0364, 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

**2021/02082**

- [the noise being echo, reverberation of the speech]

**2021/02085**

- [Periodic noise]

**2021/02087**

- [the noise being separate speech, e.g. cocktail party]

**21/0216**

- characterised by the method used for estimating noise

**2021/02161**

- [Number of inputs available containing the signal or the noise to be suppressed]

**2021/02163**

- [Only one microphone]

**2021/02165**

- [Two microphones, one receiving mainly the noise signal and the other one mainly the speech signal]

**2021/02166**

- [Microphone arrays; Beamforming]

**2021/02168**

- [the estimation exclusively taking place during speech pauses]

**21/0224**

- Processing in the time domain

**21/0232**

- Processing in the frequency domain
Time compression or expansion for improving intelligibility signals for synchronising with other signals, e.g. video by changing speed using band spreading techniques

Voice signal separating by changing the amplitude

Voice signal separating using properties of sound source

Automatic adjustment for improving intelligibility

G10L 21/0308

Details of processing therefor

Automatic adjustment

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

Details of processing therefor

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.

2021/03643 ... [Diver speech]

WARNING

Group G10L 2021/0364 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.

2021/03646 ... [Stress or Lombard effect]

WARNING

Group G10L 2021/0364 is incomplete pending reclassification of documents from group G10L 21/0202.

Groups G10L 21/0202 and G10L 2021/0364 should be considered in order to perform a complete search.

Transformation of speech into a non-audible representation, e.g. speech visualisation or speech processing for tactile aids (G10L 15/26 takes precedence)

(Aids for the handicapped in understanding)

Transforming into visible information

{Synthesis of the lips movements from speech, e.g. for talking heads}

by displaying time domain information

by displaying frequency domain information

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)

Details of the transformation process

Speech or voice analysis techniques not restricted to a single one of groups G10L 21/00 - G10L 21/01 (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)

is characterised by the type of extracted parameters

is characterised by the extracted parameters being correlation coefficients

is characterised by the extracted parameters being zero crossing rates

is characterised by the extracted parameters being prediction coefficients

is characterised by the extracted parameters being formant information

is characterised by the extracted parameters being spectral information of each sub-band

is characterised by the type of analysis window

specially adapted for particular use

for comparison or discrimination

... for retrieval

... for processing of video signals

... for measuring the quality of voice signals

... for estimating an emotional state

... for extracting parameters related to health condition (detecting or measuring for diagnostic purposes A61B 5/00)

... for evaluating synthetic or decoded voice signals

... for transmitting synthetic or decoded voice signals

... for modelling vocal tract parameters

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)

(based on threshold decision)

[Adaptive threshold]

... for discriminating voice from music

... for discriminating voice from noise

... Detection of discrete points within a voice signal

... Pitch determination of speech signals

... [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