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.

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 synthisisers
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}
13/034  . Details of speech synthesis systems, e.g. synthesiser structure or memory management
13/043 . {Synthesisers specially adapted to particular applications}
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}
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}
2015/083 . {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}
15/088 . . . { 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}
Constructional details of speech recognition systems

Speech to text systems (G10L 15/08 takes precedence)

Speech recognition using non-acoustical features

Procedures used during a speech recognition process, e.g. man-machine dialogue

Speaker identification or verification

Speaker recognisers specially adapted for particular applications (G07C 9/00071 takes precedence)

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

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

Score normalisation

Use of phonemic categorisation or speech recognition prior to speaker recognition or verification

Hidden Markov models [HMMs]

Artificial neural networks; Connectionist approaches

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

Interactive procedures; Man-machine interfaces

the user being prompted to utter a password or a predefined phrase

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

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)

Codebooks

Codebook adaptations

Backward prediction of gain

Design or structure of the codebook

Multi-stage vector quantisation

Tree or trellis structures; Delayed decisions

Codebook element generation

Algebraic codebooks

Orthogonal codebooks

Interpolation of codebook vectors

Long term prediction filters, i.e. pitch estimation

Smoothing of parameters of the decoder interpolation

Codebook search algorithms

Selection criteria for distances

Viterbi algorithms

Codebook for LPC parameters

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

Speech coding using phonetic or linguistic decoding of the source; Reconstruction using text-to-speech synthesis
19/0019 \[\text{Vocoder specially adapted for particular applications}\]
19/002 \ Dynamic bit allocation (for perceptual audio coders 19/002)
19/005 \ Correction of errors induced by the transmission channel, if related to the coding algorithm
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)
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\}
19/022 \ \{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\}
19/028 \ \{Noise substitution, i.e. substituting non-tonal spectral components by noisy source\}
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 \ \{Vocoders 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\}
2021/0135 \ \{Voice conversion or morphing\}
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)\}
21/0202 \ \{Applications\}
21/0205 \ \{Enhancement of intelligibility of clean or coded speech\}
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\}
21/0264 \ \{characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques\}
21/0272 . . Voice signal separating
21/028 . . using properties of sound source
21/0308 . . characterised by the type of parameter
measurement, e.g. correlation techniques, zero
crossing techniques or predictive techniques
21/0316 . . by changing the amplitude
21/0324 . . Details of processing therefor
21/0332 . . involving modification of waveforms
21/034 . . Automatic adjustment
21/0356 . . for synchronising with other signals, e.g. video
signals
21/0364 . . for improving intelligibility
2021/03643 . . . [Diver speech]
2021/0346 . . . . (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
2021/0575 . . . {Aids for the handicapped in speaking}
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)
2021/065 . . . {Aids for the handicapped in understanding}
21/10 . . Transforming into visible information
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}
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