**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**.

**Guide heading:**

<table>
<thead>
<tr>
<th>Subclass</th>
<th>Description</th>
</tr>
</thead>
</table>
| **G10L 13/00** | **Speech synthesis**  
|                | **Text to speech systems**                                                 |
| **G10L 13/02** | Methods for producing synthetic speech  
|                | Speech synthesisers                                                         |
| **G10L 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**) |
| **G10L 13/033**| Voice editing, e.g. manipulating the voice of the synthesiser               |
| **G10L 13/035**| Voice editing, e.g. manipulating the voice of the synthesiser               |
|                | { Pitch control }                                                           |
| **G10L 13/04** | Details of speech synthesis systems, e.g. synthesiser structure or memory  
|                | management                                                                 |
| **G10L 13/043**| Details of speech synthesis systems, e.g. synthesiser structure or memory  
|                | management                                                                 |
|                | { Synthesisers specially adapted to particular applications }               |

**WARNING**

This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to **G10L 13/00** and subgroups.

<table>
<thead>
<tr>
<th>Subclass</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G10L 13/047</strong></td>
<td>Architecture of speech synthesisers</td>
</tr>
</tbody>
</table>
| **G10L 13/06** | Elementary speech units used in speech synthesisers  
|                | Concatenation rules                                                         |
| **G10L 13/07** | Concatenation rules                                                         |
| **G10L 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                                                               |
| **G10L 13/086**| 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                                                               |
|                | { Detection of language }                                                    |
Prosody rules derived from text
Stress or intonation

Speech recognition (G10L 17/00 takes precedence)

Language recognition
Assessment or evaluation of speech recognition systems
Feature extraction for speech recognition
Selection of recognition unit
Segmentation
Word boundary detection
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)

Training
Adaptation to the speaker
{ supervised, i.e. under machine guidance }

Speech classification or search
Recognition networks (G10L 15/142, G10L 15/16 take precedence)
using distance or distortion measures between unknown speech and reference templates
using dynamic programming techniques, e.g. dynamic time warping [DTW]
using statistical models, e.g. hidden Markov models [HMMs] (G10L 15/18 takes precedence)

Hidden Markov Models [HMMs]
Training of HMMs
(with insufficient amount of training data, e.g. state sharing, tying, deleted interpolation)
Duration modelling in HMMs, e.g. semi HMM, segmental models or transition probabilities

using artificial neural networks
using natural language modelling
using prosody or stress
Semantic context, e.g. disambiguation of the recognition hypotheses based on word meaning
Parsing for meaning understanding
using context dependencies, e.g. language models
Phonemic context, e.g. pronunciation rules, phonotactical constraints or phoneme n-grams
Grammatical context, e.g. disambiguation of the recognition hypotheses
based on word sequence rules

G10L 15/193 . . . . . Formal grammars, e.g. finite state automata, context free grammars or word networks

G10L 15/197 . . . . . Probabilistic grammars, e.g. word n-grams

G10L 15/20 . Speech recognition techniques specially adapted for robustness in adverse environments, e.g. in noise, of stress induced speech (G10L 21/02 takes precedence)

G10L 15/22 . Procedures used during a speech recognition process, e.g. man-machine dialogue

G10L 15/222 . { Barge in, i.e. overridable guidance for interrupting prompts }

G10L 15/24 . Speech recognition using non-acoustical features

G10L 15/25 . . using position of the lips, movement of the lips or face analysis

G10L 15/26 . Speech to text systems (G10L 15/08 takes precedence)

G10L 15/265 . . { 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) }

**WARNING**

This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to G10L 15/00 and subgroups.

G10L 15/28 . Constructional details of speech recognition systems

G10L 15/285 . . (Memory allocation or algorithm optimisation to reduce hardware requirements )

G10L 15/30 . . Distributed recognition, e.g. in client-server systems, for mobile phones or network applications

G10L 15/32 . . Multiple recognisers used in sequence or in parallel

Score combination systems therefor, e.g. voting systems

G10L 15/34 . . Adaptation of a single recogniser for parallel processing, e.g. by use of multiple processors or cloud computing

G10L 17/00 Speaker identification or verification

G10L 17/005 . { Speaker recognisers specially adapted for particular applications (G07C 9/00071 takes precedence) }

**WARNING**

This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to G10L 17/00 and subgroups.

G10L 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

G10L 17/04 . Training, enrolment or model building

G10L 17/06 . Decision making techniques
   Pattern matching strategies

G10L 17/08 . Use of distortion metrics or a particular distance between probe pattern and reference templates

G10L 17/10 . Multimodal systems, i.e. based on the integration of multiple recognition engines or fusion of expert systems

G10L 17/12 . Score normalisation

G10L 17/14 . Use of phonemic categorisation or speech recognition prior to speaker recognition or verification

G10L 17/16 . Hidden Markov models [HMMs]

G10L 17/18 . Artificial neural networks
   Connectionist approaches

G10L 17/20 . Pattern transformations or operations aimed at increasing system robustness, e.g. against channel noise or different working conditions

G10L 17/22 . Interactive procedures
   Man-machine interfaces

G10L 17/24 . the user being prompted to utter a password or a predefined phrase

G10L 17/26 . Recognition of special voice characteristics, e.g. for use in lie detectors
   Recognition of animal voices

G10L 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 )

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

G10L 19/0018 . (Speech coding using phonetic or linguistical decoding of the source; Reconstruction using text-to-speech synthesis )

G10L 19/0019 . ( Vocoder specially adapted for particular applications )

WARNING
This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to G10L 19/00 and subgroups.

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

G10L 19/005 . Correction of errors induced by the transmission channel, if related to the coding
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)

Comfort noise or silence coding

Audio watermarking, i.e. embedding inaudible data in the audio signal

using spectral analysis, e.g. transform vocoders or subband vocoders

{ using subband decomposition }

{Subband vocoders }

{ using orthogonal transformation }

{using wavelet decomposition }

Blocking, i.e. grouping of samples in time
Choice of analysis windows
Overlap factoring

Detection of transients or attacks for time/frequency resolution switching

Noise substitution, i.e. substituting non-tonal spectral components by noisy source (comfort noise for discontinuous speech transmission G10L 19/012)

Spectral prediction for preventing pre-echo
Temporary noise shaping [TNS], e.g. in MPEG2 or MPEG4

Quantisation or dequantisation of spectral components

Scalar quantisation

Vector quantisation, e.g. TwinVQ audio

using predictive techniques

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 (G10L 25/90 takes precedence)

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]

Determination or coding of the excitation function being a code excited linear prediction [CELP] vocoders
Residual excited linear prediction [RELP]
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 }

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
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)
{ Applications }

WARNING
This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to G10L 19/00 and subgroups.

Noise filtering
characterised by the method used for estimating noise
Processing in the time domain
Processing in the frequency domain
characterised by the type of parameter measurement, e.g. correlation
techniques, zero crossing techniques or predictive techniques

G10L 21/0272 ... Voice signal separating
G10L 21/028 ... using properties of sound source
G10L 21/0308 ... characterised by the type of parameter measurement, e.g. correlation
techniques, zero crossing techniques or predictive techniques
G10L 21/0316 ... by changing the amplitude
G10L 21/0324 ... Details of processing therefor
G10L 21/0332 ... involving modification of waveforms
G10L 21/034 ... Automatic adjustment
G10L 21/0356 ... for synchronising with other signals, e.g. video signals
G10L 21/0364 ... for improving intelligibility
G10L 21/038 ... using band spreading techniques
G10L 21/0388 ... Details of processing therefor

G10L 21/04 ... Time compression or expansion
G10L 21/043 ... by changing speed
G10L 21/045 ... using thinning out or insertion of a waveform
G10L 21/047 ... characterised by the type of waveform to be thinned out or inserted
G10L 21/049 ... characterised by the interconnection of waveforms
G10L 21/055 ... for synchronising with other signals, e.g. video signals
G10L 21/057 ... for improving intelligibility

G10L 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)
G10L 21/10 ... transforming into visible information
G10L 21/12 ... by displaying time domain information
G10L 21/14 ... by displaying frequency domain information
G10L 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)
G10L 21/18 ... Details of the transformation process

G10L 25/00 Speech or voice analysis techniques not restricted to a single one of groups G10L
15/00-G10L 21/00

G10L 25/03 ... characterised by the type of extracted parameters
G10L 25/06 ... the extracted parameters being correlation coefficients
G10L 25/09 ... the extracted parameters being zero crossing rates
G10L 25/12 ... the extracted parameters being prediction coefficients
G10L 25/15 ... the extracted parameters being formant information
G10L 25/18 ... the extracted parameters being spectral information of each sub-band
G10L 25/21 ... the extracted parameters being power information
G10L 25/24 ... the extracted parameters being the cepstrum
G10L 25/27 ... characterised by the analysis technique
using neural networks
using fuzzy logic
using chaos theory
using genetic algorithms
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 results of analysis
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)
for discriminating voice from music
for discriminating voice from noise
Detection of discrete points within a voice signal
Pitch determination of speech signals
Discriminating between voiced and unvoiced parts of speech signals (G10L 25/90 takes precedence)
Subject matter not provided for in other groups of this subclass

Guide heading:

Speech synthesis
Text to speech systems

Methods for producing synthetic speech
Speech synthesisers
{ Overlap-add techniques }

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
{ Special characters, e.g. punctuation marks }
G10L 2013/10 . . Prosody rules derived from text
Stress or intonation

G10L 2013/105 . . { Duration }

G10L 2015/00 Speech recognition (G10L 17/00 takes precedence)

G10L 2015/02 . . Feature extraction for speech recognition
Selection of recognition unit

G10L 2015/022 . . { Demisyllables, biphones or triphones being the recognition units }
G10L 2015/025 . . { Phonemes, fenemes or fenones being the recognition units }
G10L 2015/027 . . { Syllables being the recognition units }

G10L 2015/06 . . 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)

G10L 2015/063 . . { Training }
G10L 2015/0631 . . { Creating reference templates; Clustering }
G10L 2015/0633 . . . . { using lexical or orthographic knowledge sources }
G10L 2015/0635 . . . . { updating or merging of old and new templates; Mean values; Weighting }
G10L 2015/0636 . . . . { Threshold criteria for the updating }
G10L 2015/0638 . . . . { Interactive procedures }

G10L 2015/08 . . Speech classification or search
G10L 2015/081 . . { Search algorithms, e.g. Baum-Welch or Viterbi }
G10L 2015/085 . . { Methods for reducing search complexity, pruning }
G10L 2015/086 . . { Recognition of spelled words }
G10L 2015/088 . . { Word spotting }

G10L 2015/22 . . Procedures used during a speech recognition process, e.g. man-machine dialogue
G10L 2015/221 . . Announcement of recognition results
G10L 2015/223 . . Execution procedure of a spoken command
G10L 2015/225 . . Feedback of the input speech
G10L 2015/226 . . Taking into account non-speech characteristics
G10L 2015/227 . . . . of the speaker
Human-factor methodology
G10L 2015/228 . . . . of application context

G10L 2019/00 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 )

G10L 2019/0001 . . { Codebooks }
G10L 2019/0002 . . { Codebook adaptations }
G10L 2019/0003 . . { Backward prediction of gain }
G10L 2019/0004 . . { Design or structure of the codebook }
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
- Adapting to target pitch
- Voice conversion or morphing

- 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)
- Noise filtering
- The noise being echo, reverberation of the speech
- Periodic noise
- The noise being separate speech, e.g. cocktail party
- Characterised by the method used for estimating noise
- Number of inputs available containing the signal or the noise to be suppressed
- Only one microphone
- Two microphones, one receiving mainly the noise signal and the other mainly the speech signal
- Microphone arrays; Beamforming
- The estimation exclusively taking place during speech pauses
- By changing the amplitude
- For improving intelligibility
- Diver speech
- Stress or Lombard effect
- Time compression or expansion
- For improving intelligibility
G10L 2021/0575    . . . { Aids for the handicapped in speaking }

G10L 2021/06    . Transformation of speech into a non-audible representation, e.g. speech visualisation or speech processing for tactile aids (G10L 15/26 takes precedence)

G10L 2021/065    . . { Aids for the handicapped in understanding }

G10L 2021/10    . . transforming into visible information

G10L 2021/105    . . . { Synthesis of the lips movements from speech, e.g. for talking heads }

G10L 2025/00    Speech or voice analysis techniques not restricted to a single one of groups G10L 15/00-G10L 21/00

G10L 2025/78    . 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)

G10L 2025/783    . . { based on threshold decision }

G10L 2025/786    . . . { Adaptive threshold }

G10L 2025/90    . Pitch determination of speech signals

G10L 2025/903    . . { using a laryngograph }

G10L 2025/906    . . { Pitch tracking }

G10L 2025/93    . Discriminating between voiced and unvoiced parts of speech signals (G10L 25/90 takes precedence)

G10L 2025/932    . . { Decision in previous or following frames }

G10L 2025/935    . . { Mixed voiced class; Transitions }

G10L 2025/937    . . { Signal energy in various frequency bands }