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.

G10L 13/00 Speech synthesis; Text to speech systems
G10L 13/02 Methods for producing synthetic speech; Speech synthesisers
G10L 2013/021 "Overlap-add techniques"
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 13/0335 "Pitch control"
G10L 13/04 Details of speech synthesis systems, e.g. synthesiser structure or memory management
G10L 13/043 "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.

G10L 13/047 Architecture of speech synthesisers
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 2013/083 "Special characters, e.g. punctuation marks"
G10L 13/086 "Detection of language"
G10L 13/10 Prosody rules derived from text; Stress or intonation
G10L 2013/105 "Duration"

G10L 15/00 Speech recognition (G10L 17/00 takes precedence)
G10L 15/005 "Language recognition"
Assessment or evaluation of speech recognition systems

Feature extraction for speech recognition; Selection of recognition unit

(Demisyllables, biphones or triphones being the recognition units)

(Phonemes, fenemes or fenones being the recognition units)

(Syllables being the recognition units)

Segmentation; Word boundary detection

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)

Creating reference templates; Clustering

(using lexical or orthographic knowledge sources)

(updating or merging of old and new templates; Mean values; Weighting)

(Threshold criteria for the updating)

{Interactive procedures}

Adaptation

to the speaker

Speech classification or search

(Search algorithms, e.g. Baum-Welch or Viterbi)

(Recognition networks (G10L 15/142, G10L 15/16 take precedence))

(Methods for reducing search complexity, pruning)

(Recognition of spelled words)

(Word spotting)

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

Formal grammars, e.g. finite state automata, context free grammars or word networks

Probabilistic grammars, e.g. word n-grams

Speech recognition techniques specially adapted for robustness in adverse environments, e.g. in noise, of stress induced speech

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

Speech recognition using non-acoustical features

using position of the lips, movement of the lips or face analysis

Speech to text systems

Constructional details of speech recognition systems

Memory allocation or algorithm optimisation to reduce hardware requirements

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

Multiple recognisers used in sequence or in parallel; Score combination systems therefor, e.g. voting systems

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

Speaker identification or verification

Speaker recognisers specially adapted for particular applications

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.
This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to \text{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
\textbf{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\}

G10L 2019/0005
- \{Multi-stage vector quantisation\}

G10L 2019/0006
- \{Tree or trellis structures; Delayed decisions\}

G10L 2019/0007
- \{Codebook element generation\}

G10L 2019/0008
- \{Algebraic codebooks\}

G10L 2019/0009
- \{Orthogonal codebooks\}

G10L 2019/001
- \{Interpolation of codebook vectors\}

G10L 2019/0011
- \{Long term prediction filters, i.e. pitch estimation\}

G10L 2019/0012
- \{Smoothing of parameters of the decoder interpolation\}

G10L 2019/0013
- \{Codebook search algorithms\}

G10L 2019/0014
- \{Selection criteria for distances\}
G10L 2019/0015 . . . {Viterbi algorithms}
G10L 2019/0016 . . {Codebook for LPC parameters}
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 . {Vocoders 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 algorithm
G10L 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)
G10L 19/012 . Comfort noise or silence coding
G10L 19/018 . Audio watermarking, i.e. embedding inaudible data in the audio signal
G10L 19/02 . using spectral analysis, e.g. transform vocoders or subband vocoders
G10L 19/0204 . . {using subband decomposition}
G10L 19/0208 . . {Subband vocoders}
G10L 19/0212 . . {using orthogonal transformation}
G10L 19/0216 . . {using wavelet decomposition}
G10L 19/022 . . Blocking, i.e. grouping of samples in time; Choice of analysis windows; Overlap factoring
G10L 19/025 . . . Detection of transients or attacks for time/frequency resolution switching
G10L 19/028 . . . Noise substitution, i.e. substituting non-tonal spectral components by noisy source (comfort noise for discontinuous speech transmission G10L 19/012)
G10L 19/03 . . Spectral prediction for preventing pre-echo; Temporary noise shaping [TNS], e.g. in MPEG2 or MPEG4
G10L 19/032 . . . Quantisation or dequantisation of spectral components
G10L 19/035 . . . Scalar quantisation
G10L 19/038 . . . Vector quantisation, e.g. TwinVQ audio
G10L 19/04 . . using predictive techniques
G10L 19/06 . . Determination or coding of the spectral characteristics, e.g. of the short-term prediction coefficients
G10L 19/07 . . . Line spectrum pair [LSP] vocoders
G10L 19/08 . . . Determination or coding of the excitation function; Determination or coding of the long-term prediction parameters
G10L 19/083 . . . 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]
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
classified by the process used
Correction of time axis
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)
{Applications}
This group is no longer used for the classification of new documents as from September 1, 2012. The backlog is being reclassified to G10L 21/00 and subgroups.

G10L 21/0205 ... {Enhancement of intelligibility of clean or coded speech}

**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 21/0364, G10L 21/057.

G10L 21/0208 ... Noise filtering
G10L 2021/02082 ... {the noise being echo, reverberation of the speech}
G10L 2021/02085 ... {Periodic noise}
G10L 2021/02087 ... {the noise being separate speech, e.g. cocktail party}
G10L 21/0216 ... characterised by the method used for estimating noise
G10L 2021/02161 ... {Number of inputs available containing the signal or the noise to be suppressed}
G10L 2021/02163 ... {Only one microphone}
G10L 2021/02165 ... {Two microphones, one receiving mainly the noise signal and the other one mainly the speech signal}
G10L 2021/02166 ... {Microphone arrays; Beamforming}
G10L 2021/02168 ... {the estimation exclusively taking place during speech pauses}
G10L 21/0224 ... Processing in the time domain
G10L 21/0232 ... Processing in the frequency domain
G10L 21/0264 ... characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques
G10L 2021/0272 ... Voice signal separating
G10L 2021/028 ... using properties of sound source
G10L 2021/0308 ... 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 2021/0324 ... Details of processing therefor
G10L 2021/0332 ... involving modification of waveforms
G10L 2021/034 ... Automatic adjustment
G10L 2021/0356 ... for synchronising with other signals, e.g. video signals
G10L 2021/0364 ... for improving intelligibility
G10L 2021/03643 ... {Diver speech}
G10L 2021/03646 ... {Stress or Lombard effect}
G10L 2021/038 ... using band spreading techniques
G10L 2021/0388 ... Details of processing therefor
G10L 21/04 ... Time compression or expansion
G10L 21/043 ... by changing speed
using thinning out or insertion of a waveform

characterised by the type of waveform to be thinned out or inserted

characterised by the interconnection of waveforms

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

for improving intelligibility

{Aids for the handicapped in speaking}

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 15/00-G10L 21/00

characterised by the type of extracted parameters

the extracted parameters being correlation coefficients

the extracted parameters being zero crossing rates

the extracted parameters being prediction coefficients

the extracted parameters being formant information

the extracted parameters being spectral information of each sub-band

the extracted parameters being power information

the extracted parameters being the cepstrum

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)
G10L 25/69 . . for evaluating synthetic or decoded voice signals
G10L 25/72 . . for transmitting results of analysis
G10L 25/75 . . for modelling vocal tract parameters
G10L 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)
G10L 2025/783 . . {based on threshold decision}
G10L 2025/786 . . {Adaptive threshold}
G10L 25/81 . . for discriminating voice from music
G10L 25/84 . . for discriminating voice from noise
G10L 25/87 . . Detection of discrete points within a voice signal
G10L 25/90 . . Pitch determination of speech signals
G10L 2025/903 . . {using a laryngograph}
G10L 2025/906 . . {Pitch tracking}
G10L 25/93 . . 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}

G10L 99/00 Subject matter not provided for in other groups of this subclass