### 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 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.

<table>
<thead>
<tr>
<th>13/00</th>
<th>Speech synthesis; Text to speech systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>13/02</td>
<td>Methods for producing synthetic speech; Speech synthesisers</td>
</tr>
<tr>
<td>2013/021</td>
<td>{Overlap-add techniques}</td>
</tr>
<tr>
<td>13/027</td>
<td>Concept to speech synthesisers; Generation of natural phrases from machine-based concepts (generation of parameters for speech synthesis out of text G10L 13/08)</td>
</tr>
<tr>
<td>13/033</td>
<td>Voice editing, e.g. manipulating the voice of the synthesiser</td>
</tr>
<tr>
<td>13/0335</td>
<td>{Pitch control}</td>
</tr>
<tr>
<td>13/04</td>
<td>Details of speech synthesis systems, e.g. synthesiser structure or memory management</td>
</tr>
<tr>
<td>13/043</td>
<td>{Synthesisers specially adapted to particular applications}</td>
</tr>
<tr>
<td>13/047</td>
<td>Architecture of speech synthesisers</td>
</tr>
<tr>
<td>13/06</td>
<td>Elementary speech units used in speech synthesisers; Concatenation rules</td>
</tr>
<tr>
<td>13/07</td>
<td>Concatenation rules</td>
</tr>
<tr>
<td>13/08</td>
<td>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</td>
</tr>
<tr>
<td>2013/083</td>
<td>{Special characters, e.g. punctuation marks}</td>
</tr>
<tr>
<td>13/086</td>
<td>{Detection of language}</td>
</tr>
<tr>
<td>13/10</td>
<td>Prosody rules derived from text; Stress or intonation</td>
</tr>
<tr>
<td>2013/105</td>
<td>{Duration}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>15/00</th>
<th>Speech recognition <em>(G10L 17/00 takes precedence)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>15/005</td>
<td>{Language recognition}</td>
</tr>
<tr>
<td>15/01</td>
<td>Assessment or evaluation of speech recognition systems</td>
</tr>
<tr>
<td>15/02</td>
<td>Feature extraction for speech recognition; Selection of recognition unit</td>
</tr>
<tr>
<td>2015/022</td>
<td>{Demisyllables, biphones or triphones being the recognition units}</td>
</tr>
<tr>
<td>2015/025</td>
<td>{Phonemes, fenemes or fenones being the recognition units}</td>
</tr>
<tr>
<td>2015/027</td>
<td>{Syllables being the recognition units}</td>
</tr>
<tr>
<td>15/04</td>
<td>Segmentation; Word boundary detection</td>
</tr>
<tr>
<td>15/05</td>
<td>Word boundary detection</td>
</tr>
<tr>
<td>15/06</td>
<td>Creation of reference templates; Training of speech recognition systems, e.g. adaptation to the characteristics of the speaker's voice <em>(G10L 15/14 takes precedence)</em></td>
</tr>
<tr>
<td>15/063</td>
<td>{Training}</td>
</tr>
<tr>
<td>2015/0631</td>
<td>{Creating reference templates; Clustering}</td>
</tr>
<tr>
<td>2015/0633</td>
<td>{using lexical or orthographic knowledge sources}</td>
</tr>
<tr>
<td>2015/0635</td>
<td>{updating or merging of old and new templates; Mean values; Weighting}</td>
</tr>
<tr>
<td>2015/0636</td>
<td>{Threshold criteria for the updating}</td>
</tr>
<tr>
<td>2015/0638</td>
<td>{Interactive procedures}</td>
</tr>
<tr>
<td>15/065</td>
<td>{Adaptation}</td>
</tr>
<tr>
<td>15/07</td>
<td>to the speaker</td>
</tr>
<tr>
<td>15/075</td>
<td>{supervised, i.e. under machine guidance}</td>
</tr>
<tr>
<td>15/08</td>
<td>Speech classification or search</td>
</tr>
<tr>
<td>2015/081</td>
<td>{Search algorithms, e.g. Baum-Welch or Viterbi}</td>
</tr>
<tr>
<td>15/083</td>
<td>{Recognition networks <em>(G10L 15/142, G10L 15/16 take precedence)</em>}</td>
</tr>
<tr>
<td>2015/085</td>
<td>{Methods for reducing search complexity, pruning}</td>
</tr>
<tr>
<td>2015/086</td>
<td>{Recognition of spelled words}</td>
</tr>
<tr>
<td>2015/088</td>
<td>{Word spotting}</td>
</tr>
<tr>
<td>15/10</td>
<td>using distance or distortion measures between unknown speech and reference templates</td>
</tr>
<tr>
<td>15/12</td>
<td>using dynamic programming techniques, e.g. dynamic time warping [DTW]</td>
</tr>
<tr>
<td>15/14</td>
<td>using statistical models, e.g. Hidden Markov Models [HMMs] <em>(G10L 15/18 takes precedence)</em></td>
</tr>
<tr>
<td>15/142</td>
<td>{Hidden Markov Models [HMMs]}</td>
</tr>
<tr>
<td>15/144</td>
<td>{Training of HMMs}</td>
</tr>
</tbody>
</table>
15/146 . . . . (with insufficient amount of training data, e.g. state sharing, tying, deleted interpolation)
15/148 . . . . (Duration modelling in HMMs, e.g. semi-HMM, segmental models or transition probabilities)
15/16 . . . . using artificial neural networks
15/18 . . . . using natural language modelling
15/1807 . . . . [using prosody or stress]
15/1815 . . . . [Semantic context, e.g. disambiguation of the recognition hypotheses based on word meaning]
15/1822 . . . . [Parsing for meaning understanding]
15/183 . . . . using context dependencies, e.g. language models
15/187 . . . . [Grammatical context, e.g. disambiguation of the recognition hypotheses based on word sequence rules]
15/193 . . . . [Formal grammars, e.g. finite state automata, context free grammars or word networks]
15/197 . . . . Probabilistic grammars, e.g. word n-grams
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)
15/22 . . . . Procedures used during a speech recognition process, e.g. man-machine dialogue
2015/221 . . . . [Announcement of recognition results]
2015/222 . . . . [Barge in, i.e. overridable guidance for interrupting prompts]
2015/223 . . . . [Execution procedure of a spoken command]
2015/225 . . . . [Feedback of the input speech]
2015/226 . . . . [Taking into account non-speech characteristics]
2015/227 . . . . [of the speaker; Human-factor methodology]
2015/228 . . . . [of application context]
15/24 . . . . Speech recognition using non-acoustical features
15/25 . . . . using position of the lips, movement of the lips or face analysis
15/26 . . . . Speech to text systems (G10L 15/08 takes precedence)
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)]
15/28 . . . . Constructional details of speech recognition systems
15/285 . . . . [Memory allocation or algorithm optimisation to reduce hardware requirements]
15/30 . . . . Distributed recognition, e.g. in client-server systems, for mobile phones or network applications
15/32 . . . . Multiple recognisers used in sequence or in parallel; Score combination systems therefor, e.g. voting systems
15/34 . . . . Adaptation of a single recogniser for parallel processing, e.g. by use of multiple processors or cloud computing

17/00 Speaker identification or verification
17/005 . . . . [Speaker recognisers specially adapted for particular applications (G07C 9/00071 takes precedence)]
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
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)

19/001 . . . . [Codebooks]
19/0002 . . . . [Codebook adaptations]
19/0003 . . . . [Backward prediction of gain]
19/0004 . . . . [Design or structure of the codebook]
19/0005 . . . . [Multi-stage vector quantisation]
19/0006 . . . . [Tree or trellis structures; Delayed decisions]
19/0007 . . . . [Codebook element generation]
19/0008 . . . . [Algebraic codebooks]
19/0009 . . . . [Orthogonal codebooks]
19/001 . . . . [Interpolation of codebook vectors]
19/0011 . . . . [Long term prediction filters, i.e. pitch estimation]
19/0012 . . . . [Smoothing of parameters of the decoder interpolation]
19/0013 . . . . [Codebook search algorithms]
19/0014 . . . . [Selection criteria for distances]
19/0015 . . . . [Viterbi algorithms]
19/0016 . . . . [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 linguistic decoding of the source; Reconstruction using text-to-speech synthesis]
using predictive techniques
subband vocoders
using spectral analysis, e.g. transform vocoders or
in the audio signal
Comfort noise or silence coding
Audio watermarking, i.e. 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 25/00 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
[VSRELP]
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
[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
[Enhancement of intelligibility of clean or
coded speech]
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 one mainly
the speech signal]
[Microphone arrays; Beamforming]
[the estimation exclusively taking place
during speech pauses]
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
Speech or voice analysis techniques not restricted to a single one of groups G10L 15/00-G10L 21/00

25/00 . Audio signal processing
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 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

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