

# 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 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.

<b>13/00</b>	<b>Speech synthesis; Text to speech systems</b>	15/04	• Segmentation; Word boundary detection
13/02	• Methods for producing synthetic speech; Speech synthesisers	15/05	• • Word boundary detection
2013/021	• • {Overlap-add techniques}	15/06	• Creation of reference templates; Training of speech recognition systems, e.g. adaptation to the characteristics of the speaker's voice ( <a href="#">G10L 15/14</a> takes precedence)
13/027	• • Concept to speech synthesisers; Generation of natural phrases from machine-based concepts (generation of parameters for speech synthesis out of text <a href="#">G10L 13/08</a> )	15/063	• • {Training}
13/033	• • Voice editing, e.g. manipulating the voice of the synthesiser	2015/0631	• • • {Creating reference templates; Clustering}
13/0335	• • • {Pitch control}	2015/0633	• • • • {using lexical or orthographic knowledge sources}
13/04	• • Details of speech synthesis systems, e.g. synthesiser structure or memory management	2015/0635	• • • • {updating or merging of old and new templates; Mean values; Weighting}
13/047	• • • Architecture of speech synthesisers	2015/0636	• • • • • {Threshold criteria for the updating}
13/06	• Elementary speech units used in speech synthesisers; Concatenation rules	2015/0638	• • • • {Interactive procedures}
13/07	• • Concatenation rules	15/065	• • Adaptation
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	15/07	• • • to the speaker
2013/083	• • {Special characters, e.g. punctuation marks}	15/075	• • • • {supervised, i.e. under machine guidance}
13/086	• • {Detection of language}	15/08	• Speech classification or search
13/10	• • Prosody rules derived from text; Stress or intonation	2015/081	• • {Search algorithms, e.g. Baum-Welch or Viterbi}
2013/105	• • • {Duration}	15/083	• • {Recognition networks ( <a href="#">G10L 15/142</a> , <a href="#">G10L 15/16</a> take precedence)}
<b>15/00</b>	<b>Speech recognition (<a href="#">G10L 17/00</a> takes precedence)</b>	2015/085	• • {Methods for reducing search complexity, pruning}
15/005	• {Language recognition}	2015/086	• • {Recognition of spelled words}
15/01	• Assessment or evaluation of speech recognition systems	2015/088	• • {Word spotting}
15/02	• Feature extraction for speech recognition; Selection of recognition unit	15/10	• • using distance or distortion measures between unknown speech and reference templates
2015/022	• • {Demisyllables, biphones or triphones being the recognition units}	15/12	• • using dynamic programming techniques, e.g. dynamic time warping [DTW]
2015/025	• • {Phonemes, fenemes or fenones being the recognition units}	15/14	• • using statistical models, e.g. Hidden Markov Models [HMMs] ( <a href="#">G10L 15/18</a> takes precedence)
2015/027	• • {Syllables being the recognition units}	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}

15/148	. . . . {Duration modelling in HMMs, e.g. semi HMM, segmental models or transition probabilities}	17/10	. . Multimodal systems, i.e. based on the integration of multiple recognition engines or fusion of expert systems
15/16	. . using artificial neural networks	17/12	. . Score normalisation
15/18	. . using natural language modelling	17/14	. . Use of phonemic categorisation or speech recognition prior to speaker recognition or verification
15/1807	. . . {using prosody or stress}		
15/1815	. . . {Semantic context, e.g. disambiguation of the recognition hypotheses based on word meaning}	17/16	. Hidden Markov models [HMM]
15/1822	. . . {Parsing for meaning understanding}	17/18	. Artificial neural networks; Connectionist approaches
15/183	. . . using context dependencies, e.g. language models	17/20	. Pattern transformations or operations aimed at increasing system robustness, e.g. against channel noise or different working conditions
15/187	. . . . Phonemic context, e.g. pronunciation rules, phonotactical constraints or phoneme n-grams	17/22	. Interactive procedures; Man-machine interfaces
15/19	. . . . Grammatical context, e.g. disambiguation of the recognition hypotheses based on word sequence rules	17/24	. . the user being prompted to utter a password or a predefined phrase
15/193	. . . . . Formal grammars, e.g. finite state automata, context free grammars or word networks	17/26	. Recognition of special voice characteristics, e.g. for use in lie detectors; Recognition of animal voices
15/197	. . . . . Probabilistic grammars, e.g. word n-grams	<b>19/00</b>	<b>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)</b>
15/20	. Speech recognition techniques specially adapted for robustness in adverse environments, e.g. in noise, of stress induced speech ( <a href="#">G10L 21/02 takes precedence</a> )	2019/0001	. {Codebooks}
15/22	. Procedures used during a speech recognition process, e.g. man-machine dialogue	2019/0002	. . {Codebook adaptations}
2015/221	. . {Announcement of recognition results}	2019/0003	. . {Backward prediction of gain}
15/222	. . {Barge in, i.e. overridable guidance for interrupting prompts}	2019/0004	. . {Design or structure of the codebook}
2015/223	. . {Execution procedure of a spoken command}	2019/0005	. . . {Multi-stage vector quantisation}
2015/225	. . {Feedback of the input speech}	2019/0006	. . . {Tree or treillis structures; Delayed decisions}
2015/226	. . {using non-speech characteristics}	2019/0007	. . {Codebook element generation}
2015/227	. . . {of the speaker; Human-factor methodology}	2019/0008	. . . {Algebraic codebooks}
2015/228	. . . {of application context}	2019/0009	. . . {Orthogonal codebooks}
15/24	. Speech recognition using non-acoustical features	2019/001	. . . {Interpolation of codebook vectors}
15/25	. . using position of the lips, movement of the lips or face analysis	2019/0011	. . {Long term prediction filters, i.e. pitch estimation}
15/26	. Speech to text systems ( <a href="#">G10L 15/08 takes precedence</a> )	2019/0012	. . {Smoothing of parameters of the decoder interpolation}
15/28	. Constructional details of speech recognition systems	2019/0013	. . {Codebook search algorithms}
15/285	. . {Memory allocation or algorithm optimisation to reduce hardware requirements}	2019/0014	. . . {Selection criteria for distances}
15/30	. . Distributed recognition, e.g. in client-server systems, for mobile phones or network applications	2019/0015	. . . {Viterbi algorithms}
15/32	. . Multiple recognisers used in sequence or in parallel; Score combination systems therefor, e.g. voting systems	2019/0016	. . {Codebook for LPC parameters}
15/34	. . Adaptation of a single recogniser for parallel processing, e.g. by use of multiple processors or cloud computing	19/0017	. {Lossless audio signal coding; Perfect reconstruction of coded audio signal by transmission of coding error ( <a href="#">G10L 19/24 takes precedence</a> )}
<b>17/00</b>	<b>Speaker identification or verification</b>	19/0018	. {Speech coding using phonetic or linguistic decoding of the source; Reconstruction using text-to-speech synthesis}
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	19/002	. Dynamic bit allocation ( <a href="#">for perceptual audio coders G10L 19/032</a> )
17/04	. Training, enrolment or model building	19/005	. Correction of errors induced by the transmission channel, if related to the coding algorithm
17/06	. Decision making techniques; Pattern matching strategies	19/008	. Multichannel audio signal coding or decoding using interchannel correlation to reduce redundancy, e.g. joint-stereo, intensity-coding or matrixing
17/08	. . Use of distortion metrics or a particular distance between probe pattern and reference templates	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}	19/24	. . . . Variable rate codecs, e.g. for generating different qualities using a scalable representation such as hierarchical encoding or layered encoding
19/0216	. . . {using wavelet decomposition}	19/26	. . Pre-filtering or post-filtering
19/022	. . Blocking, i.e. grouping of samples in time; Choice of analysis windows; Overlap factoring	19/265	. . . {Pre-filtering, e.g. high frequency emphasis prior to encoding}
19/025	. . . Detection of transients or attacks for time/frequency resolution switching	<b>21/00</b>	<b>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)</b>
19/028	. . Noise substitution, i.e. substituting non-tonal spectral components by noisy source (comfort noise for discontinuous speech transmission G10L 19/012)	21/003	. Changing voice quality, e.g. pitch or formants
19/03	. . Spectral prediction for preventing pre-echo; Temporary noise shaping [TNS], e.g. in MPEG2 or MPEG4	21/007	. . characterised by the process used
19/032	. . Quantisation or dequantisation of spectral components	21/01	. . . Correction of time axis
19/035	. . . Scalar quantisation	21/013	. . . Adapting to target pitch
19/038	. . . Vector quantisation, e.g. TwinVQ audio	2021/0135	. . . . {Voice conversion or morphing}
19/04	. using predictive techniques	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)
19/06	. . Determination or coding of the spectral characteristics, e.g. of the short-term prediction coefficients	21/0208	. . Noise filtering
19/07	. . . Line spectrum pair [LSP] vocoders	2021/02082	. . . {the noise being echo, reverberation of the speech}
19/08	. . Determination or coding of the excitation function; Determination or coding of the long-term prediction parameters	2021/02085	. . . {Periodic noise}
19/083	. . . the excitation function being an excitation gain (G10L 25/90 takes precedence)	2021/02087	. . . {the noise being separate speech, e.g. cocktail party}
19/087	. . . using mixed excitation models, e.g. MELP, MBE, split band LPC or HVXC	21/0216	. . . characterised by the method used for estimating noise
19/09	. . . Long term prediction, i.e. removing periodical redundancies, e.g. by using adaptive codebook or pitch predictor	2021/02161	. . . . {Number of inputs available containing the signal or the noise to be suppressed}
19/093	. . . using sinusoidal excitation models	2021/02163	. . . . . {Only one microphone}
19/097	. . . using prototype waveform decomposition or prototype waveform interpolative [PWI] coders	2021/02165	. . . . . {Two microphones, one receiving mainly the noise signal and the other one mainly the speech signal}
19/10	. . . the excitation function being a multipulse excitation	2021/02166	. . . . . {Microphone arrays; Beamforming}
19/107	. . . . Sparse pulse excitation, e.g. by using algebraic codebook	2021/02168	. . . . . {the estimation exclusively taking place during speech pauses}
19/113	. . . . Regular pulse excitation	21/0224	. . . . Processing in the time domain
19/12	. . . the excitation function being a code excitation, e.g. in code excited linear prediction [CELP] vocoders	21/0232	. . . . Processing in the frequency domain
19/125	. . . . Pitch excitation, e.g. pitch synchronous innovation CELP [PSI-CELP]	21/0264	. . . characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques
19/13	. . . . Residual excited linear prediction [RELP]	21/0272	. . Voice signal separating
19/135	. . . . Vector sum excited linear prediction [VSELP]	21/028	. . . using properties of sound source
19/16	. . Vocoder architecture	21/0308	. . . characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques
19/167	. . . {Audio streaming, i.e. formatting and decoding of an encoded audio signal representation into a data stream for transmission or storage purposes}	21/0316	. . by changing the amplitude
19/173	. . . {Transcoding, i.e. converting between two coded representations avoiding cascaded coding-decoding}	21/0324	. . . Details of processing therefor
19/18	. . . Vocoder architecture	21/0332	. . . . involving modification of waveforms
19/20	. . . . using sound class specific coding, hybrid encoders or object based coding	21/034	. . . . Automatic adjustment
19/22	. . . . Mode decision, i.e. based on audio signal content versus external parameters	21/0356	. . . for synchronising with other signals, e.g. video signals
		21/0364	. . . for improving intelligibility
		2021/03643	. . . . {Diver speech}
		2021/03646	. . . . {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	25/81	. . for discriminating voice from music
21/055	. . for synchronising with other signals, e.g. video signals	25/84	. . for discriminating voice from noise
21/057	. . for improving intelligibility	25/87	. . Detection of discrete points within a voice signal
2021/0575	. . . {Aids for the handicapped in speaking}	25/90	. Pitch determination of speech signals
21/06	. Transformation of speech into a non-audible representation, e.g. speech visualisation or speech processing for tactile aids ( <a href="#">G10L 15/26 takes precedence</a> )	2025/903	. . {using a laryngograph}
2021/065	. . {Aids for the handicapped in understanding}	2025/906	. . {Pitch tracking}
21/10	. . Transforming into visible information	25/93	. Discriminating between voiced and unvoiced parts of speech signals ( <a href="#">G10L 25/90 takes precedence</a> )
2021/105	. . . {Synthesis of the lips movements from speech, e.g. for talking heads}	2025/932	. . {Decision in previous or following frames}
21/12	. . . by displaying time domain information	2025/935	. . {Mixed voiced class; Transitions}
21/14	. . . by displaying frequency domain information	2025/937	. . {Signal energy in various frequency bands}
21/16	. . Transforming into a non-visible representation (devices or methods enabling ear patients to replace direct auditory perception by another kind of perception <a href="#">A61F 11/04</a> )		
21/18	. . Details of the transformation process		
<b>25/00</b>	<b>Speech or voice analysis techniques not restricted to a single one of groups <a href="#">G10L 15/00</a> - <a href="#">G10L 21/00</a> (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, <a href="#">H03G 3/34</a>)</b>	<b>99/00</b>	<b>Subject matter not provided for in other groups of this subclass</b>
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 ( <a href="#">detecting or measuring for diagnostic purposes A61B 5/00</a> )		
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 ( <a href="#">switching of direction of transmission by voice frequency in two-way loud-speaking telephone systems H04M 9/10</a> )		
2025/783	. . {based on threshold decision}		
2025/786	. . . {Adaptive threshold}		