

## SECTION G — PHYSICS

## G10 MUSICAL INSTRUMENTS; ACOUSTICS

## G10L SPEECH ANALYSIS OR SYNTHESIS; SPEECH RECOGNITION; SPEECH OR VOICE PROCESSING; SPEECH OR AUDIO CODING OR DECODING [4]

Note(s) [2010.01]

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.

|              |  |              |  |
|--------------|--|--------------|--|
| <b>13/00</b> | <b>Speech synthesis; Text to speech systems [7, 2006.01]</b>   | 15/14        | • • using statistical models, e.g. Hidden Markov Models [HMM] (G10L 15/18 takes precedence) [7, 2006.01]   |
| 13/02        | • Methods for producing synthetic speech; Speech synthesisers [7, 2006.01, 2013.01]  |              |  |
| 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) [2013.01]                     | 15/16        | • • using artificial neural networks [7, 2006.01]  |
|              |  | 15/18        | • • using natural language modelling [7, 2006.01, 2013.01]   |
| 13/033       | • • Voice editing, e.g. manipulating the voice of the synthesiser [2013.01]  | 15/183       | • • • using context dependencies, e.g. language models [2013.01]   |
| 13/04        | • • Details of speech synthesis systems, e.g. synthesiser structure or memory management [7, 2006.01, 2013.01]   | 15/187       | • • • • Phonemic context, e.g. pronunciation rules, phonotactical constraints or phoneme n-grams [2013.01]   |
| 13/047       | • • • Architecture of speech synthesisers [2013.01]  | 15/19        | • • • • Grammatical context, e.g. disambiguation of recognition hypotheses based on word sequence rules [2013.01]  |
| 13/06        | • Elementary speech units used in speech synthesisers; Concatenation rules [7, 2006.01, 2013.01]   | 15/193       | • • • • • Formal grammars, e.g. finite state automata, context free grammars or word networks [2013.01]  |
| 13/07        | • • Concatenation rules [2013.01]  |              |  |
| 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 [7, 2006.01, 2013.01] | 15/197       | • • • • • Probabilistic grammars, e.g. word n-grams [2013.01]  |
| 13/10        | • • Prosody rules derived from text; Stress or intonation [2013.01]  | 15/20        | • Speech recognition techniques specially adapted for robustness in adverse environments, e.g. in noise or of stress induced speech (G10L 21/02 takes precedence) [7, 2006.01] |
| <b>15/00</b> | <b>Speech recognition (G10L 17/00 takes precedence) [7, 2006.01, 2013.01]</b>  | 15/22        | • Procedures used during a speech recognition process, e.g. man-machine dialog [7, 2006.01]  |
| 15/01        | • Assessment or evaluation of speech recognition systems [2013.01]   | 15/24        | • Speech recognition using non-acoustical features [7, 2006.01, 2013.01]   |
| 15/02        | • Feature extraction for speech recognition; Selection of recognition unit [7, 2006.01]  | 15/25        | • • using position of the lips, movement of the lips or face analysis [2013.01]  |
| 15/04        | • Segmentation; Word boundary detection [7, 2006.01, 2013.01]  | 15/26        | • Speech to text systems (G10L 15/08 takes precedence) [7, 2006.01]  |
| 15/05        | • • Word boundary detection [2013.01]  | 15/28        | • Constructional details of speech recognition systems [7, 2006.01, 2013.01]   |
| 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) [7, 2006.01, 2013.01]       | 15/30        | • • Distributed recognition, e.g. in client-server systems, for mobile phones or network applications [2013.01]  |
| 15/065       | • • Adaptation [2013.01]   | 15/32        | • • Multiple recognisers used in sequence or in parallel; Score combination systems therefor, e.g. voting systems [2013.01]  |
| 15/07        | • • • to the speaker [2013.01]   |              |  |
| 15/08        | • Speech classification or search [7, 2006.01]   | 15/34        | • • Adaptation of a single recogniser for parallel processing, e.g. by use of multiple processors or cloud computing [2013.01]   |
| 15/10        | • • using distance or distortion measures between unknown speech and reference templates [7, 2006.01]  |              |  |
| 15/12        | • • using dynamic programming techniques, e.g. dynamic time warping [DTW] [7, 2006.01]   | <b>17/00</b> | <b>Speaker identification or verification [7, 2006.01, 2013.01]</b>  |

- 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 [2013.01]
- 17/04 • Training, enrolment or model building [2013.01]
- 17/06 • Decision making techniques; Pattern matching strategies [2013.01]
- 17/08 • • Use of distortion metrics or a particular distance between probe pattern and reference templates [2013.01]
- 17/10 • • Multimodal systems, i.e. based on the integration of multiple recognition engines or fusion of expert systems [2013.01]
- 17/12 • • Score normalisation [2013.01]
- 17/14 • • Use of phonemic categorisation or speech recognition prior to speaker recognition or verification [2013.01]
- 17/16 • Hidden Markov models [HMMs] [2013.01]
- 17/18 • Artificial neural networks; Connectionist approaches [2013.01]
- 17/20 • Pattern transformations or operations aimed at increasing system robustness, e.g. against channel noise or different working conditions [2013.01]
- 17/22 • Interactive procedures; Man-machine interfaces [2013.01]
- 17/24 • • the user being prompted to utter a password or a predefined phrase [2013.01]
- 17/26 • Recognition of special voice characteristics, e.g. for use in lie detectors; Recognition of animal voices [2013.01]
- 19/00 Speech or audio signal 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) [7, 2006.01, 2013.01]**
- 19/002 • Dynamic bit allocation (for perceptual audio coders G10L 19/032) [2013.01]
- 19/005 • Correction of errors induced by the transmission channel, if related to the coding algorithm [2013.01]
- 19/008 • Multichannel audio signal coding or decoding using interchannel correlation to reduce redundancy, e.g. joint-stereo, intensity-coding or matrixing [2013.01]
- 19/012 • Comfort noise or silence coding [2013.01]
- 19/018 • Audio watermarking, i.e. embedding inaudible data in the audio signal [2013.01]
- 19/02 • using spectral analysis, e.g. transform vocoders or subband vocoders [7, 2006.01, 2013.01]
- 19/022 • • Blocking, i.e. grouping of samples in time; Choice of analysis windows; Overlap factoring [2013.01]
- 19/025 • • • Detection of transients or attacks for time/frequency resolution switching [2013.01]
- 19/028 • • Noise substitution, e.g. substituting non-tonal spectral components by noisy source (comfort noise for discontinuous speech transmission G10L 19/012) [2013.01]
- 19/03 • • Spectral prediction for preventing pre-echo; Temporary noise shaping [TNS], e.g. in MPEG2 or MPEG4 [2013.01]
- 19/032 • • Quantisation or dequantisation of spectral components [2013.01]
- 19/035 • • • Scalar quantisation [2013.01]
- 19/038 • • • Vector quantisation, e.g. TwinVQ audio [2013.01]
- 19/04 • using predictive techniques [7, 2006.01, 2013.01]
- 19/06 • • Determination or coding of the spectral characteristics, e.g. of the short-term prediction coefficients [7, 2006.01, 2013.01]
- 19/07 • • • Line spectrum pair [LSP] vocoders [2013.01]
- 19/08 • • Determination or coding of the excitation function; Determination or coding of the long-term prediction parameters [7, 2006.01, 2013.01]
- 19/083 • • • the excitation function being an excitation gain (G10L 25/90 takes precedence) [2013.01]
- 19/087 • • • using mixed excitation models, e.g. MELP, MBE, split band LPC or HVXC [2013.01]
- 19/09 • • • Long term prediction, i.e. removing periodical redundancies, e.g. by using adaptive codebook or pitch predictor [2013.01]
- 19/093 • • • using sinusoidal excitation models [2013.01]
- 19/097 • • • using prototype waveform decomposition or prototype waveform interpolative [PWI] coders [2013.01]
- 19/10 • • • the excitation function being a multipulse excitation [7, 2006.01, 2013.01]
- 19/107 • • • • Sparse pulse excitation, e.g. by using algebraic codebook [2013.01]
- 19/113 • • • • Regular pulse excitation [2013.01]
- 19/12 • • • the excitation function being a code excitation, e.g. in code excited linear prediction [CELP] vocoders [7, 2006.01, 2013.01]
- 19/125 • • • • Pitch excitation, e.g. pitch synchronous innovation CELP [PSI-CELP] [2013.01]
- 19/13 • • • • Residual excited linear prediction [RELP] [2013.01]
- 19/135 • • • • Vector sum excited linear prediction [VSELP] [2013.01]
- 19/16 • • Vocoder architecture [2013.01]
- 19/18 • • • Vocoders using multiple modes [2013.01]
- 19/20 • • • • using sound class specific coding, hybrid encoders or object based coding [2013.01]
- 19/22 • • • • Mode decision, i.e. based on audio signal content versus external parameters [2013.01]
- 19/24 • • • • Variable rate codecs, e.g. for generating different qualities using a scalable representation such as hierarchical encoding or layered encoding [2013.01]
- 19/26 • • Pre-filtering or post-filtering [2013.01]
- 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) [7, 2006.01, 2013.01]**
- 21/003 • Changing voice quality, e.g. pitch or formants [2013.01]
- 21/007 • • characterised by the process used [2013.01]
- 21/01 • • • Correction of time axis [2013.01]
- 21/013 • • • Adapting to target pitch [2013.01]
- 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) [7, 2006.01, 2013.01]
- 21/0208 • • Noise filtering [2013.01]
- 21/0216 • • • characterised by the method used for estimating noise [2013.01]
- 21/0224 • • • • Processing in the time domain [2013.01]
- 21/0232 • • • • Processing in the frequency domain [2013.01]

|         |  |       |  |
|---------|--|-------|--|
| 21/0264 | • • • characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques [2013.01]   | 25/03 | • characterised by the type of extracted parameters [2013.01]  |
| 21/0272 | • • • Voice signal separating [2013.01]  | 25/06 | • • the extracted parameters being correlation coefficients [2013.01]  |
| 21/028  | • • • using properties of sound source [2013.01]   | 25/09 | • • the extracted parameters being zero crossing rates [2013.01]   |
| 21/0308 | • • • characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques [2013.01]   | 25/12 | • • the extracted parameters being prediction coefficients [2013.01]   |
| 21/0316 | • • • by changing the amplitude [2013.01]  | 25/15 | • • the extracted parameters being formant information [2013.01]   |
| 21/0324 | • • • Details of processing therefor [2013.01]   | 25/18 | • • the extracted parameters being spectral information of each sub-band [2013.01]   |
| 21/0332 | • • • • involving modification of waveforms [2013.01]  | 25/21 | • • the extracted parameters being power information [2013.01]   |
| 21/034  | • • • • Automatic adjustment [2013.01]   | 25/24 | • • the extracted parameters being the cepstrum [2013.01]  |
| 21/0356 | • • • for synchronising with other signals, e.g. video signals [2013.01]   | 25/27 | • characterised by the analysis technique [2013.01]  |
| 21/0364 | • • • for improving intelligibility [2013.01]  | 25/30 | • • using neural networks [2013.01]  |
| 21/038  | • • using band spreading techniques [2013.01]  | 25/33 | • • using fuzzy logic [2013.01]  |
| 21/0388 | • • • Details of processing therefor [2013.01]   | 25/36 | • • using chaos theory [2013.01]   |
| 21/04   | • Time compression or expansion [7, 2006.01, 2013.01]  | 25/39 | • • using genetic algorithms [2013.01]   |
| 21/043  | • • by changing speed [2013.01]  | 25/45 | • characterised by the type of analysis window [2013.01]   |
| 21/045  | • • • using thinning out or insertion of a waveform [2013.01]  | 25/48 | • specially adapted for particular use [2013.01]   |
| 21/047  | • • • • characterised by the type of waveform to be thinned out or inserted [2013.01]  | 25/51 | • • for comparison or discrimination [2013.01]   |
| 21/049  | • • • • characterised by the interconnection of waveforms [2013.01]  | 25/54 | • • • for retrieval [2013.01]  |
| 21/055  | • • for synchronising with other signals, e.g. video signals [2013.01]   | 25/57 | • • • for processing of video signals [2013.01]  |
| 21/057  | • • for improving intelligibility [2013.01]  | 25/60 | • • • for measuring the quality of voice signals [2013.01]   |
| 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) [7, 2006.01, 2013.01]  | 25/63 | • • • for estimating an emotional state [2013.01]  |
| 21/10   | • • Transforming into visible information [2013.01]  | 25/66 | • • • for extracting parameters related to health condition (detecting or measuring for diagnostic purposes A61B 5/00) [2013.01]   |
| 21/12   | • • • by displaying time domain information [2013.01]  | 25/69 | • • for evaluating synthetic or decoded voice signals [2013.01]  |
| 21/14   | • • • by displaying frequency domain information [2013.01]   | 25/72 | • • for transmitting results of analysis [2013.01]   |
| 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) [2013.01]   | 25/75 | • for modelling vocal tract parameters [2013.01]   |
| 21/18   | • • Details of the transformation process [2013.01]  | 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) [2013.01] |
| 25/00   | <b>Speech or voice analysis techniques not restricted to a single one of groups G10L 15/00-G10L 21/00</b> (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, H03G 3/34) [2013.01] | 25/81 | • • for discriminating voice from music [2013.01]  |
|         |  | 25/84 | • • for discriminating voice from noise [2013.01]  |
|         |  | 25/87 | • • Detection of discrete points within a voice signal [2013.01]   |
|         |  | 25/90 | • Pitch determination of speech signals [2013.01]  |
|         |  | 25/93 | • Discriminating between voiced and unvoiced parts of speech signals (G10L 25/90 takes precedence) [2013.01]   |
|         |  | 99/00 | <b>Subject matter not provided for in other groups of this subclass [2013.01]</b>  |