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.

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.
Constructional details of speech recognition systems (G10L 15/08)

Speech recognition using non-acoustical features (G10L 15/08)

Pattern transformations or operations aimed at increasing system robustness, e.g. against channel noise or different working conditions (G10L 15/08)

Interactive procedures; Man-machine interfaces (G10L 15/08)

the user being prompted to utter a password or a predefined phrase (G10L 15/08)

Recognition of special voice characteristics, e.g. for use in lie detectors; Recognition of animal voices (G10L 15/08)

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)

Speaker identification or verification (G10L 17/00)

Speaker recognisers specially adapted for particular applications (G07C 9/25 takes precedence) (G10L 17/00)

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 17/00)

Training, enrolment or model building (G10L 17/00)

Decision making techniques; Pattern matching strategies (G10L 17/00)

Use of distortion metrics or a particular distance between probe pattern and reference templates (G10L 17/00)

Multimodal systems, i.e. based on the integration of multiple recognition engines or fusion of expert systems (G10L 17/00)

Score normalisation (G10L 17/00)

Use of phonemic categorisation or speech recognition prior to speaker recognition or verification (G10L 17/00)

Hidden Markov models [HMMs] (G10L 17/00)

Artificial neural networks; Connectionist approaches (G10L 17/00)

Pattern transformations or operations aimed at increasing system robustness, e.g. against channel noise or different working conditions (G10L 17/00)

Interactive procedures; Man-machine interfaces (G10L 17/00)

the user being prompted to utter a password or a predefined phrase (G10L 17/00)

Recognition of special voice characteristics, e.g. for use in lie detectors; Recognition of animal voices (G10L 17/00)
19/0204 . . . [using subband decomposition]
19/0208 . . . [Subband vocoders]
19/0212 . . . [using orthogonal transformation]
19/0216 . . . [using wavelet decomposition]
19/022 . . . Blocking, i.e. grouping of samples in time; Choice of analysis windows; Overlap fatiguing
19/025 . . . Detection of transients or attacks for time/ frequency resolution switching
19/028 . . . Noise substitution, i.e. substituting non-tonal spectral components by noisy source (comfort noise for discontinuous speech transmission G10L 19/012)
19/03 . . . Spectral prediction for preventing pre-echo; Temporary noise shaping [TNS], e.g. in MPEG2 or MPEG4
19/032 . . . Quantisation or dequantisation of spectral components
19/035 . . . Scalar quantisation
19/038 . . . Vector quantisation, e.g. TwinVQ audio
19/04 . . . using predictive techniques
19/06 . . . Determination or coding of the spectral characteristics, e.g. of the short-term prediction coefficients
19/07 . . . Line spectrum pair [LSP] vocoders
19/08 . . . Determination or coding of the excitation function; Determination or coding of the long-term prediction parameters
19/083 . . . the excitation function being an excitation gain (G10L 25/90 takes precedence)
19/087 . . . using mixed excitation models, e.g. MELP, MBE, split band LPC or HVXC
19/09 . . . Long term prediction, i.e. removing periodical redundancies, e.g. by using adaptive codebook or pitch predictor
19/093 . . . using sinusoidal excitation models
19/097 . . . using prototype waveform decomposition or prototype waveform interpolative [PWI] coders
19/10 . . . the excitation function being a multipulse excitation
19/107 . . . Sparse pulse excitation, e.g. by using algebraic codebook
19/113 . . . Regular pulse excitation
19/12 . . . the excitation function being a code excitation, e.g. in code excited linear prediction [CELP] vocoders
19/125 . . . Pitch excitation, e.g. pitch synchronous innovation CELP [PSI-CELP]
19/13 . . . Residual excited linear prediction [RELP]
19/135 . . . . . . Vector sum excited linear prediction [VSELP]
19/16 . . . Vocoder architecture
19/167 . . . . . . [Audio streaming, i.e. formatting and decoding of an encoded audio signal representation into a data stream for transmission or storage purposes]
19/173 . . . . . . [Transcoding, i.e. converting between two coded representations avoiding cascaded coding-decoding]
19/18 . . . Vocoders using multiple modes
19/20 . . . using sound class specific coding, hybrid encoders or object based coding
19/22 . . . . Mode decision, i.e. based on audio signal content versus external parameters
19/24 . . . Variable rate codecs, e.g. for generating different qualities using a scalable representation such as hierarchical encoding or layered encoding
19/26 . . . Pre-filtering or post-filtering
19/265 . . . . [Pre-filtering, e.g. high frequency emphasis prior to encoding]

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)
21/003 . . . Changing voice quality, e.g. pitch or formants
21/007 . . . characterised by the process used
21/01 . . . . Correction of time axis
21/013 . . . Adapting to target pitch
21/018 . . . . [Voice conversion or morphing]
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)
21/0202 . . . [Applications]
21/0205 . . . [Enhancement of intelligibility of clean or coded speech]
21/0208 . . . Noise filtering
21/021 . . . . . . (the noise being echo, reverberation of the speech)
21/02082 . . . . . . (the noise being separate speech, e.g. cocktail party)
21/021 . . . . . . characterised by the method used for estimating noise
21/02161 . . . . . . [Number of inputs available containing the signal or the noise to be suppressed]
21/02163 . . . . . . [Only one microphone]
21/02165 . . . . . . [Two microphones, one receiving mainly the noise signal and the other one mainly the speech signal]
21/02166 . . . . . . [Microphone arrays; Beamforming]
21/02168 . . . . . . [The estimation exclusively taking place during speech pauses]
21/0224 . . . . Processing in the time domain
21/0232 . . . . Processing in the frequency domain
21/026 . . . . characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques
21/0272 . . . . Voice signal separating
21/028 . . . . using properties of sound source
characterised by the type of parameter measurement, e.g. correlation techniques, zero crossing techniques or predictive techniques

by changing the amplitude

Details of processing therefor involving modification of waveforms

Automatic adjustment

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

for improving intelligibility

{ Diver speech }

{ Stress or Lombard effect }

using band spreading techniques

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

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

using neural networks

using fuzzy logic

using chaos theory

using genetic algorithms

specially adapted for particular use

for comparison or discrimination