G10L

SPEECH ANALYSIS OR SYNTHESIS; SPEECH RECOGNITION ; SPEECH OR VOICE PROCESSING; SPEECH OR AUDIO CODING OR DECODING

Definition statement

This subclass/group covers:

- processing of speech or voice signals in general (G10L 25/00);
- production of synthetic speech signals (G10L 13/00);
- recognition of speech (G10L 15/00);
- lyrics recognition from a singing voice (G10L 15/00);
- speaker identification, authentication or verification (G10L 17/00);
- singer recognition from a singing voice (G10L 17/00);
- analysis of speech signals for bandwidth compression or extension, bit-rate or redundancy reduction (G10L 19/00);
- coding/decoding of audio signals for compression and expansion using analysis-synthesis, source filter models or psycho-acoustic analysis (G10L 19/00);
- modification of speech signals, speech enhancement, source separation (G10L 21/00);
- noise filtering or echo cancellation in an audio signal (G10L 21/00);
- speech or voice analysis techniques specially adapted to analyse or modify audio signals not necessarily including speech or voice are also covered in subgroups (G10L 21/00,G10L 25/00);

References relevant to classification in this subclass

This subclass/group does not cover:

Devices for the storage of speech signals	G11B
Spatial sound recording	H04R 5/00
Spatial sound reproduction	H04S
Encoding of compressed speech signals for transmission or storage	H04L

Coding or synthesis of audio signals in musical instruments	G10H
Karaoke or singing voice processing	G10H
Sound production	G10K
Sound input or sound output arrangements for computers	G06F 3/16
Amplifiers	H03F
Gain or frequency control	H03G 3/00
Broadcasting	H04H
Secret communication	H04K 1/00
Handling natural language data	G06F 17/20
General pattern recognition	G06K 9/00
Speech or voice prosthesis	A61F 2/20
Mere application of speech or voice analysis techniques	application place

Examples of places where the subject matter of this group is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Information retrieval of audio data	G06F 17/3074
Broadcasting arrangements of audio	H04H 60/58
Name dialling controlled by voice recognition	H04M 1/271
Automatic arrangements for answering calls	H04M 1/64

Places in relation to which this subclass is residual:

Acoustics not otherwise provided for	G10K 15/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Measurement of sound waves in general	G01H
Sound input/output for computers	G06F 3/00
Image data processing	G06T
Teaching or communicating with the blind, deaf or mute	G09B
Electronic musical instruments	G10H
Information storage, e.g. sound storage	G11B
Electronic circuits for sound generation	H03B
Electronic filters	H03H
Coding, decoding or code conversion, error protection in general	H03M
Telephonic communication	H04M
Switching systems	H04Q
Microphone arrangements, hearing aids, public address systems	H04R
Spatial sound reproduction	H04S

Glossary of terms

In this subclass/group, the following terms (or expressions) are used with the meaning indicated:

In this subclass, the following terms are used with the meaning indicated:

Speech	definite vocal sounds that form words to express thoughts and ideas
Voice	sounds generated by vocal chords or synthetic versions thereof
Audio	of or relating to humanly audible sound

G10L 13/00

Speech synthesis; Text to speech systems

Definition statement

This subclass/group covers:

- synthesis of speech from text, concatenation of smaller speech units, grapheme to phoneme conversion;
- modification of the voice for speech synthesis: gender, age, pitch, prosody, stress.
- hardware or software implementation details of a speech synthesis system

References relevant to classification in this group

This subclass/group does not cover:

Examples of places where the subject matter of this group is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Navigation systems for vehicles, guidance using speech synthesis.	G01C 21/3629
Speech synthesis in games	A63F 9/24
Electric switches with speech feedback	H01H
Speech synthesis in mobile phones	H04M 1/00
Electronic musical instruments	G10H

Sound producing other than musical instruments	G10K

Informative references

Attention is drawn to the following places, which may be of interest for search:

Natural language translation	G06F 17/28
Excitation coding of a speech signal	G10L 19/08

Synonyms and Keywords

In patent documents the following abbreviations are often used:

HMM	Hidden Markov Model
TTS	Text To Speech

G10L 13/027

Concept-to-speech synthesisers; Generation of natural phrases not from text but from machine-based concepts

Definition statement

This subclass/group covers:

concepts used for speech synthesis can be linked to an emotion to be conveyed (US2010329505), a communication goal driving a dialogue (US2010241420), image-to-speech (US2010231752), native sounding speech (US2004030554)

References relevant to classification in this group

This subclass/group does not cover:

Language translation	G06F 17/28

G10L 15/00

Speech recognition (G10L17/00 takes precedence)

Definition statement

This subclass/group covers:

- recognition of text or phonemes from a spoken audio signal;
- spoken dialog interfaces, human-machine spoken interfaces
- topic detection in a dialogue, semantic analysis, keyword detection, spoken command and control
- context dependent speech recognition (location, environment, age, gender, etc.)
- parameter extraction, acoustic models, word models, grammars, language models for speech recognition
- recognition of speech in a noisy environment
- recognition of speech using visual clues
- feedback of the recognition results, disambiguation of speech recognition results
- dedicated hardware or software implementations, parallel and distributed processing of speech recognition engines

References relevant to classification in this group

This subclass/group does not cover:

Examples of places where the subject matter of this group is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Voice control for systems within a vehicle	B60R 16/0373
Speech input for vehicle navigation systems	G01C 21/3608
Sound input arrangements for computers	G06F 3/16
Teaching how to speak	G09B 19/04
Name dialling controlled by voice recognition	H04M 1/271

Speech interaction details in automatic or semi-automatic exchange systems for interactive information services	H04M 3/4936
Spoken command and control of surgical instruments	A61B 17/00
Speech input in video games	A63F 13/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Complex mathematical functions	G06F 17/10
Handling natural language data	G06F 17/20
Information retrieval of audio data	G06F 17/3074
Educational appliances	G09B 5/06
Face recognition, lip reading without acoustical input	G06K 9/00221
Pattern recognition	G06K 9/00
Signal processing for recording	G11B 20/00
Natural language processing	G06F 17/20

Synonyms and Keywords

In patent documents the following abbreviations are often used:

ANN	Artificial neural network
ASR	Automatic speech recognition
CSR	Continuous speech recognition
GMM	Gaussian mixture model

HMM	Hidden Markov model
IVR	Interactive voice response
MLP	Multi layer perceptron
VLSR	Very large speech recognition

G10L 17/00

Speaker identification or verification

Definition statement

This subclass/group covers:

- recognition, identification of a speaker
- verification, authentication of a speaker
- feature extraction, dialog, prompts, passwords for identification
- identification in noisy condition
- multimodal identification including voice
- impostor detection

Informative references

Attention is drawn to the following places, which may be of interest for search:

Complex mathematical functions	G06F 17/10
Information retrieval of audio data	G06F 17/3074
Secret secure communication including means for verifying the identity or authority of a user	H04L 9/32
Security arrangements, restricting access by authenticating users, using biometric data	G06F 21/00N5A2B
Pattern recognition	G06K 9/00
Individual entry or exit registers,	G07C 9/00071

access control with identity check using personal physical data	

Glossary of terms

In this subclass/group, the following terms (or expressions) are used with the meaning indicated:

In this subclass, the following terms are used with the meaning indicated:

Speaker verification, or authentication	refers to verifying that the user claimed identity is real, he is otherwise an impostor. Speaker recognition, or identification, aims at determining who the user is among a closed (finite number) set of users. He is otherwise unknown.
A goat, sheep	often refers to a person whose voice is easy to counterfeit.
A wolf, predator	often refers to a person who can easily counterfeit someone else's voice or is often identified as someone else.
An impostor	is someone actively trying to counterfeit someone else's identity.

Synonyms and Keywords

In patent documents the following abbreviations are often used:

ANN	Artificial neural network
ASR	Automatic speech recognition
GMM	Gaussian mixture model
HMM	Hidden Markov model
IVR	Interactive voice response
MLP	Multi layer perceptron

G10L 19/00

Speech or audio signal analysis-synthesis techniques for redundancy reduction, e.g. in vocoders ; Coding or decoding of speech or audio signals, e.g. for compression or expansion, source-filter models or psychoacoustic analysis (in electric musical instruments G10H)

Definition statement

This subclass/group covers:

Techniques for the reduction of data from audio sources, i.e. compression of audio. These techniques are applied to reduce the quantity of information to be stored or transmitted, but are independent of the end-application, medium or transmission channel, i.e. do only exploit the properties of the source signal itself or the final receiver exposed to this signal (the listener).

Mainly two types of sources can be distinguished :

"speech only" encompass signals produced by human speakers, and historically was to be understood as mono-channel, single speaker "telephone quality" speech having a narrow bandwidth limited to max. 4kHz. Encoding of speech only sources primarily aim at reducing the bit-rate while still providing fair intelligibility of the spoken content, but not always fidelity to the original.

"Audio signal" is broader and comprises speech as well as background information, e.g. music source having multiple channels. Encoding of audio deals primarily with transparent, i.e. "high fidelity" reproduction of the original signal.

The compression techniques can also be distinguished as being :

Lossy or Lossless, i.e. whether a perfect reconstruction of the source is possible, or only a perceptually acceptable approximation can be done.

The techniques classified in this subclass are based either on modelling the production of the signal (voice) or the perception of it (general audio).

References relevant to classification in this group

This subclass/group does not cover:

Coding of signals within electronic musical instruments	G10H

Informative references

Attention is drawn to the following places, which may be of interest for search:

Complex mathematical functions	G06F 17/10
Quality monitoring in automatic, semi automatic exchanges	H04M 3/2236
Quality control of voice transmission between switching centres	H04M 7/00M18
Signal processing not specific to the method of recording or reproducing	G11B 20/00
Editing; Indexing; Addressing; Timing or synchronizing; Monitoring;	G11B 27/00
Compression	H03M 7/30
Detecting, preventing errors in received information	H04L 1/00
Transmission of audio and video in television systems	H04N 7/52A
Simultaneous speech and data transmission	H04M 11/06
Stereophonic arrangements	H04R 5/00
Stereophonic systems	H04S
Wireless communication networks	H04W

Glossary of terms

In this subclass/group, the following terms (or expressions) are used with the meaning indicated:

In this subclass, the following terms are used with the meaning indicated:

audio signal	is meant to include speech, music, silence or background signal, or any combinations thereof, unless explicitly specified

Synonyms and Keywords

In patent documents the following abbreviations are often used:

CELP	Code Excited Linear Prediction
CTX	Continuous transmission
DTX	Discontinuous transmission
HVXC	Harmonic Vector eXcitation Coding
LPC	linear prediction coding
MBE	Multiband Excitation
MELP	Mixed Excitation Linear Prediction
MOS	mean opinion score
MPEG	Moving Picture Experts Group
MPEG1 audio	Standard ISO/IEC 11172-3
MPEG2 audio	Standard ISO/IEC 13818-3
MPEG4 audio	Standard ISO/IEC 14496-3
MP3	MPEG 1 Layer III
PCM	pulse code modulation
PWI	Prototype Waveform Interpolation
SBR	Spectral Band Replication

In patent documents the following expressions/words " perceptual" and "psychoacoustic" are often used as synonyms.

G10L 19/24

Variable rate or variable quality codecs, e.g. scalable representation, hierarchical or layered encoding

Definition statement

This subclass/group covers:

Coding of a signal with rate adaptation, e.g. adapted to voiced speech, unvoiced speech, transitions and noise/silence portions.

Coding of a signal with a core encoder providing a minimum level of quality, and extension layers to improve the quality but requiring a higher bitrate. It includes parameter based bandwidth extension (i.e. SBR) or channel extension.

This group is in opposition to G10L 21/038 in which the bandwidth extension is artificial, i.e. based on the only narrowband encoded signal.

Informative references

Attention is drawn to the following places, which may be of interest for search:

Artificial bandwidth extension, i.e. based on the only narrowband encoded signal	G10L 21/038
Spatial sound recording	H04R 5/00
Spatial sound reproduction	H04S

G10L 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 (G10L19/00 takes precedence)

Definition statement

This subclass/group covers:

The subgroup deals with speech or voice modification applications, but receives also applications for speech or voice analysis techniques specially adapted to analyse or modify audio signals not necessarily including speech or voice but which are not music signals (G10H).

- bandwidth extension of an audio signal
- improvement of the intelligibility of a coded speech signal

- removal of noise from an audio signal
- removal of echo from an audio signal
- separation of audio sources
- pitch, speed modification of an audio signal
- voice morphing
- visualisation of audio signals (e.g. sonagrams)
- lips or face movement synchronisation with speech (e.g phonemes - visemes alignment).
- face animation synchronisation with the emotion contained in the voice or speech signal

References relevant to classification in this group

This subclass/group does not cover:

Speech or audio signal analysis-synthesis techniques for redundancy reduction, e.g. in vocoders ; Coding or decoding of speech or audio signals, e.g. for compression or expansion, source-filter models or psychoacoustic analysis	G10L 19/00

Electronic musical instruments	G10H
Loudspekers, microphones, gramophone pick-up or like acoustic electromechanical transducers; deaf-aid sets; public address systems	H04R
Stereophonic systems	H04S

Informative references

Attention is drawn to the following places, which may be of interest for search:

Complex mathematical functions	G06F 17/10
Signal processing not specific to the	G11B 20/00

method of recording or reproducing,	
Gain control in amplifiers	H03G 3/32
Animation based on audio data, talking heads	G06T 13/205
Signal processing not specific to the method of recording or reproducing, for reducing noise	G11B 20/24
Editing; Indexing; Addressing; Timing or synchronizing; Monitoring;	G11B 27/00
Direction finder	G01S 3/00
Reducing noise or bandwidth in transmission systems not characterised by the medium used for transmission	H04B 14/046
Reducing echo effect or singing in line transmissions systems	H04B 3/20
Hearing aids	H04R 25/00
Public address systems	H04R 27/00

Glossary of terms

In this subclass/group, the following terms (or expressions) are used with the meaning indicated:

In this subclass, the following terms are used with the meaning indicated:

Viseme	a visual representation of the mouth, lips, tongue and teeth corresponding to a phoneme

Synonyms and Keywords

In patent documents the following abbreviations are often used:

BSS	blind source separation

LDA	linear discriminant analysis
NB	narrowband
PCA	principal component analysis
SBR	Spectral Band Replication
WB	wideband

G10L 21/0356

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

Definition statement

This subclass/group covers:

Visemes are selected to match with the corresponding speech segment, or the speech segments are adapted/chosen, to match with the viseme. This symbol also encompasses the coarticulation effects as used in facial character animation or talking heads.

Informative references

Attention is drawn to the following places, which may be of interest for search:

Facial character animation per se	G06T 13/205

G10L 21/038

using band spreading

Definition statement

This subclass/group covers:

Bandwidth extension taking place at the receiving side, e.g. generation of artificial low or high frequency components, regeneration of spectral holes, based on the only narrowband encoded signal. This is in opposition with G10L 19/24 wherein parameters are computed during the encoding step to enable bandwidth extension at the decoding step.

Informative references

Attention is drawn to the following places, which may be of interest for search:

Parameter based bandwidth extension (e.g. SBR)	G10L 19/24

G10L 25/00

Speech or voice analysis techniques not restricted to a single one of groups G10L15/00-G10L21/00

Definition statement

This subclass/group covers:

- processing of speech or voice signals in general, in particular detection of a speech signal, end points detection in noise, extraction of pitch, measure of the voicing, emotional state, voice pathology or other speech or voice related parameters

- speech or voice analysis techniques specially adapted to analyse audio signals not necessarily including speech or voice, such as audio scene segmentation, jingle detection, separation from music or noise, detection of particular sounds;

References relevant to classification in this group

This subclass/group does not cover:

Karaoke or singing voice processing, parameter extraction for musical signal categorisation, electronic musical instruments	G10H
Gain or frequency control	H03G 3/342
DTX communication	H04J 3/175
Multiplex systems	H04Q 1/30

Informative references

Attention is drawn to the following places, which may be of interest for search:

Switching of direction of transmission by voice in loud-speaking telephone systems	H04M 9/10

Comfort noise	G10L 19/00N

Glossary of terms

In this subclass/group, the following terms (or expressions) are used with the meaning indicated:

In this subclass, the following terms are used with the meaning indicated:

audio signal	is of or relating to humanly audible sound. e.g., it comprises any combination of background noise or silence, voice or speech, music.

G10L 99/00

Subject matter not provided for in other groups of this subclass