BESCHWERDEKAMMERN
DES EUROPÄISCHEN
PATENTAMTS

BOARDS OF APPEAL OF
THE EUROPEAN PATENT
OFFICE

CHAMBRES DE RECOURS
DE L'OFFICE EUROPÉEN
DES BREVETS

Internal distribution code:
(A) [-] Publication in OJ
(B) [-] To Chairmen and Members
(C) [-] To Chairmen
(D) [X] No distribution

Datasheet for the decision
of 7 April 2017

Case Number: T 2650/11 - 3.4.01

Application Number: 07872094.3

Publication Number: 2135240

IPC: G10L19/00

Language of the proceedings: EN

Title of invention:
SPEECH CODING SYSTEM AND METHOD

Applicant:
Skype

Headword:

Relevant legal provisions:
EPC 1973 Art. 54(1), 54(2), 56

Keyword:
Inventive step - main request (no)
Novelty - auxiliary request (no)

Decisions cited:

Catchword:

D E C I S I O N
of Technical Board of Appeal 3.4.01
of 7 April 2017

Appellant:	Skype
(Applicant)	70 Sir John Rogerson's Quay
	Dublin 2 (IE)
Representative:	Driver, Virginia Rozanne
	Page White & Farrer
	Bedford House
	John Street
	London WC1N 2BF (GB)
Decision under appeal:	**Decision of the Examining Division of the European Patent Office posted on 30 June 2011 refusing European patent application No. 07872094.3 pursuant to Article 97(2) EPC.**

Composition of the Board:

Chairman	G. Assi
Members:	P. Fontenay
	J. Geschwind

Summary of Facts and Submissions

I.   The examining division refused European patent
        application No. 07 872 094.

        The examining division held that the subject-matter of
        independent claims of a main request and first and
        second auxiliary requests underlying the decision was
        not new in the sense of Art. 54(1),(2) EPC 1973. In
        this respect, the examining division relied on prior
        art document:

        (D9) S. van de Par et al., "*Scalable Noise Coder for
              Parametric Sound Coding*", Audio Engineering
              Society, 118th Convention on 28-31 May 2005 in
              Barcelona (ES), Vol. 118, Convention Paper 6465,
              pages 1-8).

        The examining division further held that the subject-
        matter of independent claims according to a third
        auxiliary request underlying the decision was not
        inventive in the sense of Art. 56 EPC with regard to
        D9, considered to illustrate the closest prior art, and
        document:

        (D8) US-A-2003/233234.

II.   The appellant (applicant) filed an appeal against the
        decision to refuse the application.

III.   The appellant requested that the decision under appeal
        be set aside and that a patent be granted on the basis
        of a set of claims according to a main request or an
        auxiliary request, as filed with the statement setting
        out the of appeal.

IV.   In accordance with an appellant's request, a summons to attend oral proceedings was issued.

V.   In a communication pursuant to Art. 15(1) RPBA, the appellant was informed of the provisional opinion of the Board with regard to the requests then pending.

        Concerning claim 1 of the main request, the recited feature "*zero additional bits for generating the artificially generated noise signal are encoded in the encoded signal*" was considered to be devoid of any technical meaning and could thus not define any clear distinguishing feature with respect to the prior art, in particular document D9.

        With regard to claim 1 of the auxiliary request, the feature "*said noise signal being placed at the at least one harmonic such that the noise signal tapers off from the peak of the at least one harmonic toward a spectral valley between harmonics*", it was observed that sais feature was not disclosed in D9.

VI.   In a letter of reply dated 21 March 2017, the appellant reversed the order of the requests filed with the grounds of appeal.

VII.   Oral proceedings before the Board took place as scheduled in the absence of the appellant's representative, as announced.

VIII.   Claim 1 of the appellant's main request reads:

        "*1.   A system (300) for enhancing a signal regenerated from an encoded speech signal (302) encoded with a model-based harmonic sinusoidal speech encoder, the system comprising:*

a harmonic sinusoidal speech decoder (304) arranged to receive the encoded speech signal and produce a decoded speech signal (306) including a voiced speech signal;

a feature extraction means (308) arranged to receive at least one of the decoded and encoded speech signal and extract at least one feature from at least one of the decoded and encoded speech signal;

a mapping means (310) arranged to map said at least one feature to an artificially generated noise signal (312) and operable to generate and output said noise signal;

characterised in that:

said noise signal has a frequency band that is within the decoded speech signal frequency band;

said system further comprises a mixing means (320) arranged to receive said decoded speech signal and said noise signal and mix said noise signal with the voiced speech signal in the decoded speech signal frequency band; and

said mixing means is further arranged to receive said voiced speech signal, to determine a location of at least one harmonic from said voiced speech signal, and to adapt the mixing of said noise signal with said voiced speech signal in dependence on the location of the at least one harmonic determined by the mixing means, said noise signal being placed at the least [sic] one harmonic such that the noise signal tapers off from the peak of the at least one harmonic toward a spectral valley between harmonics."

Claims 2 to 19 of the main request depend on claim 1.

Claim 20 of the main request reads:

"*20.   A method of enhancing a signal regenerated from an encoded speech signal (302) encoded with a model-based harmonic sinusoidal speech encoder, the method comprising:*

receiving the encoded speech signal at a terminal;

producing a decoded speech signal (306) including voiced frames;

extracting at least one feature from at least one of the decoded and encoded speech signal;

mapping said at least one feature to an artificially generated noise signal (312) and generating said noise signal; and

mixing said noise signal and the voiced frames of said decoded speech signal;

characterised in that:

said noise signal has a frequency band that is within the decoded speech signal frequency band; and

said mixing comprises receiving said voiced speech signal, determining a location of at least one harmonic from said voiced speech signal, and adapting the mixing of said noise signal with said voiced speech signal in dependence on said the [sic] determined location of the at least one harmonic, said noise signal being placed at the one harmonic such that the noise signal tapers off from the peak of the at least one harmonic toward a spectral valley between harmonics."

IX.   Claim 1 of the appellant's auxiliary request reads:

"*1. A system (300) comprising a destination terminal for enhancing a signal regenerated from an encoded speech signal (302), the destination terminal comprising:*

a decoder (304) arranged to receive the encoded speech signal and produce a decoded speech signal (306) comprising a voiced speech signal;

a feature extraction means (308) arranged to receive at least one of the decoded and encoded speech signal and extract at least one feature from at least one of the decoded and encoded speech signal; and

a mapping means (310) arranged to map said at least one feature to an artificially generated noise signal (312) and operable to generate and output said noise signal;

characterised in that:

the artificially generated noise signal has a frequency band that is within the decoded speech signal frequency band, and said system further comprises a mixing means (320) arranged to receive said decoded speech signal and said noise signal and mix said noise signal with the voiced speech signal in the decoded speech signal frequency band, the mixer thereby being arranged to mix the noise signal at a location in the spectrum of the decoded speech signal having a received power at that location in the spectrum; and

zero additional bits for generating the artificially generated noise signal are encoded in the encoded signal, and instead the at least one feature extracted from at least one of the decoded and encoded speech signal is used to provide information about how to generate said noise signal at the receiving terminal, said at least one feature including at least one of: a fundamental frequency, a location of each harmonic in a sinusoidal description, and a harmonic amplitude and phase."

Claims 2 to 21 of the auxiliary request depend on claim 1.

Claim 22 of the auxiliary request reads:

"*22.   A method of enhancing a signal regenerated from an encoded speech signal (302) comprising:*

receiving the encoded speech signal at a destination terminal;

producing a decoded speech signal (306) comprising a voiced speech signal;

extracting at least one feature from at least one of the decoded and encoded speech signal;

mapping said at least one feature to an artificially generated noise signal (312) and generating said noise signal,; and

mixing said noise signal and the voiced speech signal of said decoded speech signal;

characterised in that:

the artificially generated noise signal has a frequency band that is within the decoded speech signal frequency band, said noise signal being mixed at a location in the spectrum of the decoded speech signal having a received power at that location in the spectrum; and

zero additional bits for generating the artificially generated noise signal are encoded in the encoded signal, and instead the at least one feature extracted from at least one of the decoded and encoded speech signal is used to provide information about how to generate said noise signal at the receiving terminal, said at least one feature including at least one of: a fundamental frequency, a location of each harmonic in a sinusoidal description, and a harmonic amplitude and phase."

Reasons for the Decision

1.   The appeal is admissible.

2.   *Main request – Art. 56 EPC 1973*

2.1   Document D9 relates to scalable noise coders for
        parametric sound coding. It thus belongs to the same
        technical field as the present invention. D9 further
        shares a common purpose with the claimed subject-matter
        in that it describes solutions to the problems
        resulting from limited bit rates in the transmission of
        audio data (cf. abstract). Moreover, D9 suggests
        tackling this issue by making use of noise modelling
        techniques (cf. page 2, left column, lines 33-36).

        For these reasons, document D9 is considered to
        illustrate the closest prior art.

2.2   The two-part form adopted for claim 1 of the main
        request is based on document D9. As acknowledged by the
        applicant, the features of the preamble are known from
        D9. Concretely, D9 discloses a system (cf. Figure 1)
        comprising a destination terminal for enhancing a
        signal regenerated from an encoded speech signal. The
        destination terminal comprises a decoder (cf. Figure 1,
        lower part) arranged to receive the encoded speech
        signal and produce a decoded speech signal comprising a
        voiced speech signal, a feature extraction means
        (sinusoidal decoder) being arranged to receive at least
        one of the decoded and encoded speech signal and
        extract at least one feature from at least one of the
        decoded and encoded speech signal. The system of D9
        further comprises a mapping means (noise modeler)
        arranged to map said at least one feature to an
        artificially generated noise signal, operable to
        generate and output said noise signal (cf. page 5, left
        column, third paragraph).

2.3   In the appellant's opinion, document D9 did not
         disclose the characterising features of claim 1
         according to which:

         - said noise signal has a frequency band that is
            within the decoded speech signal frequency band,

         - said system comprises a mixing means (320) arranged
            to receive said decoded speech signal and said noise
            signal and mix said noise signal with the voiced
            speech signal in the decoded speech signal frequency
            band, and

         - said mixing means is further arranged to receive
            said voiced speech signal, to determine a location
            of at least one harmonic from said voiced speech
            signal, and to adapt the mixing of said noise signal
            with said voiced speech signal in dependence on the
            location of the at least one harmonic determined by
            the mixing means, said noise signal being placed at
            the at least one harmonic such that the noise signal
            tapers off from the peak of the at least one
            harmonic toward a spectral valley between harmonics.

2.3.1   As described in D9 (cf. section 2.2, in particular page
         5, right-hand column, last paragraph, page 6, left-hand
         column), the iterative process disclosed therein leads
         to the determination of various energy values X_i for
         each predetermined noise band and, thus, indirectly to
         the determination of the noise signal W. This
         conclusion results directly from step 2 of said process
         in which X_i is determined by reference to the various
         excitation pattern for each i, that is, for each sub-
         band, of the original signal (cf. equation 8 in section
         2.2). It might indeed be that the iterative process
         leads, depending on the situation, to certain values of

X_i ($X_i \neq 0$) for bands that are not represented within the decoded speech signal. However, equation (9) in D9 clearly discloses that the artificially generated signal, as a whole, overlaps with the decoded speech signal. In other terms, the artificially generated signal is well within the decoded speech signal frequency band, contrary to the appellant's view.

2.3.2   Moreover, as illustrated in Figure 1 of D9, the decoder comprises a mixing means arranged to receive the decoded speech signal (output of Sinusoidal decoder) and the noise signal (output of Noise modeler) and mix said noise signal with the voiced speech signal in the decoded speech signal frequency band as recited in the claim (cf. Figure 1, addition of the signals provided in the upper and lower branches of the decoder).

2.3.3   As it results from the presence of the sinusoidal decoder in the decoder of Figure 1, the system of D9 also discloses to determine the location of some harmonics of the signal. In the context of D9, these harmonics correspond to the main harmonics of the signals that have been encoded and transmitted, that is, the harmonics that belong to the upper layers that have not been dropped (cf. page 2, left column, lines 11-32; page 6, right column, lines 9-26).

2.4   It follows that the claimed system differs from the system disclosed in D9 in that the noise signal is placed at the at least one harmonic such that the noise signal tapers off from the peak of the at least one harmonic toward a spectral valley between harmonics.

         This distinguishing feature permits to increase the perceived naturalness and quality of the speech (cf. published application, page 2, lines 18-22; page 10,

lines 8-11; page 12, lines 31-33). In this respect, the findings according to which said quality would primarily be affected by "*metallic sounding*" artefacts would not justify a reduction of the objective problem to this specific effect. It is namely stressed that the final purpose of the invention is to increase the intelligibility of the signal transmitted independently of the perceived impression created by the decoding operation.

2.5   Document D8 is concerned with improving the perceived quality of audio signals obtained from audio coding systems (cf. D8, paragraphs [0001] and [0010]). Its teaching would have therefore been considered by the skilled person. The mere fact that D8 is primarily concerned with filling spectral gaps or holes in the received energy spectrum is no obstacle for said teaching being taken into account insofar as the skilled person would have recognised that the solutions proposed are very general and not solely limited to decoded signals with gaps in the energy spectrum.
In this respect, the question to be answered is whether the skilled person would have recognised that the teaching of D8 may be incorporated in the system of D9.

It is noted that the scaling envelopes disclosed in D8 (cf. Figures 8 and 10) could also be applied in the context of D9 to the decoded harmonics in a way similar to the one illustrated in Figure 11 of D8. It is stressed, in this respect, that the algorithm for synthesizing a noise signal, as disclosed in detail in D9 (cf. page 5, left-hand column, line 41 to page 6, left-hand column), does not constitute the sole approach which could be envisaged. On the contrary, the statement immediately preceding the discussion of the algorithm in D9 (cf. page 5, left-hand column, lines

32-40) according to which "*Algorithms for synthesizing a noise signal such that the noise part plus sinusoidal part have an excitation pattern that is very similar to that of the original signal can be made in various ways*" constitutes a clear hint to envisage alternative approaches.

In D8, the scaling envelopes are centered on the frequencies defining the limits of each gap. The scaling envelopes disclosed both with regard to Figures 8 and 10 have in common to decrease sidewards with respect to a spectral component. Applied to the decoded harmonics of D9, the teaching of D8 would thus lead to noise signals with frequencies decreasing in amplitude around said decoded harmonics, that is, to noise signals tapering off from the peak of at least one harmonic toward a spectral valley between harmonics, as recited in claim 1 of the main request.

2.6   In conclusion, the subject-matter of claim 1 according to the main request does not involve an inventive step having regard to the combination of documents D9 and D8. The same applies to the method claim 20 *mutatis mutandis*.

For these reasons, the main request is not allowable.

3.   *Auxiliary request – Art. 54(1),(2) EPC 1973*

3.1   Claim 1 of the auxiliary request differs from claim 1 of the main request, firstly, in that the distinguishing feature identified above, according to which the noise signal is placed at the at least one harmonic such that it tapers off from the peak of said harmonic toward a spectral valley between harmonics, has been deleted.

Claim 1 of the auxiliary request further differs from claim 1 of the main request in that it recites that:

- the mixer is arranged to mix the noise signal at a location in the spectrum of the decoded speech signal having a received power at that location in the spectrum, and

- zero additional bits for generating the artificially generated noise signal are encoded in the encoded signal, and instead the at least one feature extracted from at least one of the decoded and encoded speech signal is used to provide information about how to generate said noise signal at the receiving terminal, said at least one feature including at least one of a fundamental frequency, a location of each harmonic in a sinusoidal description, and a harmonic amplitude and phase.

3.2   For the reasons developed above with regard to the main request, the iterative process disclosed in D9 eventually leads to a step of mixing signal noise, as it results from the determination of the complete array of X_i parameters, at locations within the spectrum of the decoded speech signal deprived of power as well as locations having a received power (cf. equation 8), as required by the claim's wording.

3.3   The appellant held that D9 did not disclose the feature recited in claim 1 of zero additional bits for generating the artificially generated noise signal being encoded in the encoded signal.

Leaving aside the fact that the reference to additional bits does not permit to identify with regard to which

data the term "*additional*" refers, it is noted that the declared purpose of the method and system disclosed in D9 consists precisely of determining the noise signal at the decoder side without the need to send any information about the adaptation of the noise coder in the bit stream (cf. D9, Abstract).

3.4   The appellant further held that D9 did not disclose the amendment recited in claim 1 concerning the at least one feature being extracted from the encoded or decoded signal and including at least one of a fundamental frequency, a location of each harmonic in a sinusoidal description and a harmonic amplitude and phase.

        According to D9 (cf. page 3, section 2.1), however, the first layer transmitted by the encoder and later decoded by the decoder contains the excitation patterns and the most relevant sinusoidal components. Moreover, the second layer transmitted by the encoder and received by the decoder contains the next most relevant sinusoidal components (cf. D9, section 2.1, "G*eneral Outline of a Scalable Parametric Codec*"; section 3, "*Results*"). In other terms, D9 discloses the identification of the fundamental harmonics contained in the signal initially encoded, that is, their amplitudes and location within the complete spectrum.

3.5   All the features recited in claim 1 of the auxiliary request are therefore known from D9. Consequently, the subject-matter of claim 1 is not new in the sense of Art. 54(1),(2) EPC 1973. The same applies *mutatis mutandis* to the method of claim 22.

        For these reasons, the auxiliary request is not allowable.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar: The Chairman:

R. Schumacher   G. Assi

Decision electronically authenticated