Patents Act 1977 **Opinion Number** **20/21**

OPINION UNDER SECTION 74A

Patent	EP 2102619 B1
Proprietor(s)	VoiceAge Corporation
Exclusive Licensee	
Requester	Jenner & Block London LLP
Observer(s)	VoiceAge Corporation (represented by Mathys & Squire LLP)
Date Opinion issued	10 January 2022

The request

1. The comptroller has been requested to issue an opinion on the validity of claims 1 and 10 of European Patent (UK) 2102619 B1 (the patent) in the light of the following prior publications:

   "Variable Bit-Rate Coding of Speech with Phonetic Classification" by E. Paksoy, K.Srinivasan, A. Gersho, published in September/October 1994 ("**PAKSOY**")

   "CELP with priority to critical segments" by L.M. da Silva and A. Alcaim, published 11.09.1998 ("**DA SILVA**")

   "Low Rate Speech Coding using a Glottal Pulse Codebook" by Jan Lindén, Jan Skoglund, and Per Hedelin, published in October 1995 ("**LINDEN**")

Observations

2. Observations were received on 18 November 2021. Observations in reply were received on 1 December 2021 including a copy of a preliminary opinion dated 14 October 2021 issued by the German Federal Patent Court regarding the patentability of the patent and a copy of an Expert Opinion on the validity of the patent dated 14 October 2021 by Professor Peter Kabal, Professor Emeritus in the Department of Electrical & Computing Engineering of McGill University in Montreal.

3. I consider that the Expert Opinion submitted with the observations in reply, which Expert Opinion itself refers to further publications that were not submitted with the original request, is being used to further the request rather than show flaws in the observations. Therefore, I will not consider this Expert Opinion. It is fundamental to the opinions process that the requester raises their best argument at the outset. This allows the observer an opportunity to comment on the entire argument. If argument

is introduced at the observations-in-reply stage, then the observer is not able to respond and so such argument is not allowed.

4. However, I will consider the preliminary opinion of the German Federal Patent Court (discussed further below). Although it would have been preferable for this document to be submitted with the original request, it is implausible that the patent proprietor was unaware of these proceedings (I note that, in related Opinion 21/21, the patent proprietor themselves bring to light a related preliminary opinion of the German Federal Patent Court with their observations).

Matters to be considered by this Opinion

5. Section 74A of the Patents Act provides for the procedure where the Comptroller can issue, on request, non-binding opinions on questions of validity relating to novelty and inventive step.

6. Section 74A(3) of the Patents Act 1977 states:

> *The comptroller shall issue an opinion if requested to do so under subsection (1) above, but shall not do so –*
>
> *(a) in such circumstances as may be prescribed, or*
>
> *(b) if for any reason he considers it inappropriate in all the circumstances to do so.*

7. Rule 94(1)(b) of the Patents Rules 2007 provides that:

> *The comptroller shall not issue an opinion if the question upon which the opinion is sought appears to him to have been sufficiently considered in any relevant proceedings.*

8. Relevant proceedings are defined in Rule 92 as proceedings (whether pending or concluded) before the comptroller, the court, or the European Patent Office.

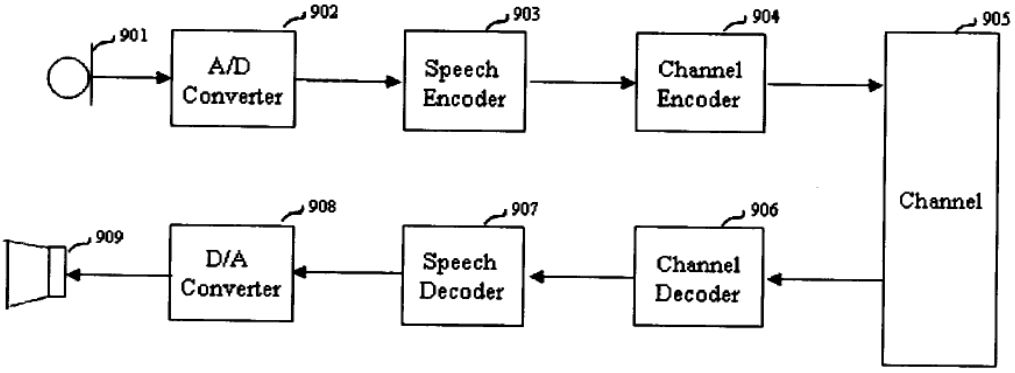
9. The German Federal Patent Court has already issued a preliminary opinion on the patentability of the patent. In this preliminary opinion, the Federal Patent Court addressed two of the three prior publications mentioned in the request, namely PAKSOY and DE SILVA. Although, it can be argued that these proceedings are not "*relevant* proceedings" under Rule 94(1)(b) as further defined by Rule 92, I have reviewed the preliminary opinion of the Federal Patent Court and I have indicated in this opinion where it appears to me that the question of validity of the patent in the light of these documents has already been answered.

The patent

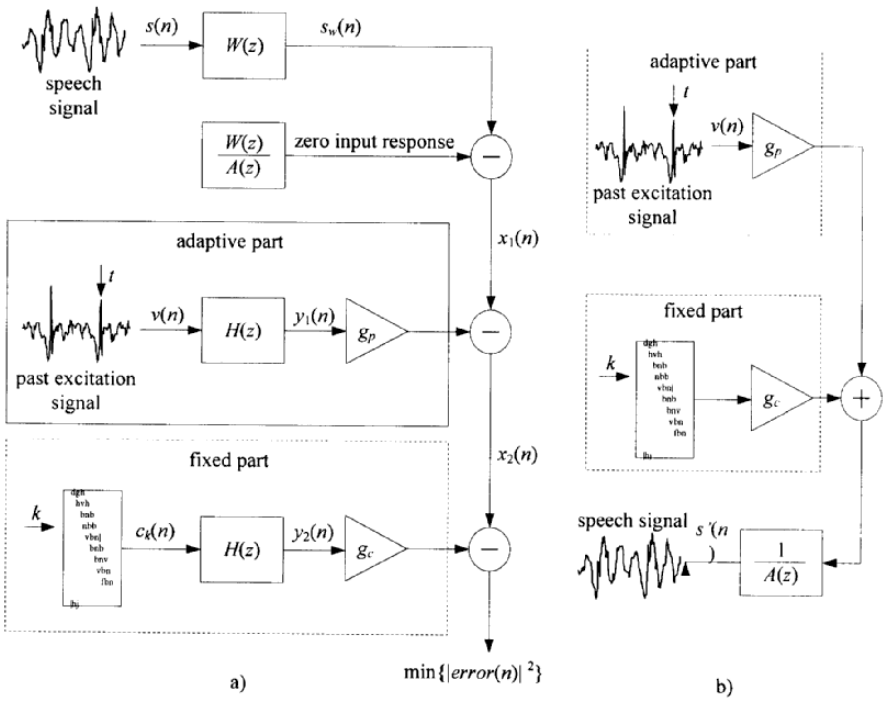
10. The patent is entitled "Method and device for coding transition frames in speech signals" and was filed on 24 October 2007 with an earlier declared priority date of 24 October 2006. The patent was granted on 22 March 2017 and remains in force in the UK.

11. The patent relates to the field of speech encoding. As explained in the patent, and

illustrated in the reproduced figure below, a speech encoder 903 converts a speech signal into a digital bit stream that is transmitted over a communication channel 905 (or stored in a storage medium). The speech signal is digitized 902 (sampled and quantized with usually 16-bits per sample) and the speech encoder 903 has the role of representing these digital samples with a smaller number of bits while maintaining a good subjective speech quality. The speech decoder 907 or synthesizer operates on the transmitted or stored bit stream and converts it back to a sound signal.



12. Code-Excited Linear Prediction (CELP) coding is one of the best prior art techniques for achieving a good compromise between subjective quality and bit rate (see figure reproduced below).



13. CELP-type speech codecs rely heavily on prediction to achieve their high performance. The prediction used can be of different kinds but usually comprises the use of an adaptive codebook containing an excitation signal selected in past frames. A CELP encoder exploits the quasi periodicity of voiced speech signal by searching in the past excitation the segment most similar to the segment being currently encoded. The same past excitation signal is maintained also in the decoder. It is then sufficient for the encoder to send a delay parameter and a gain for the decoder to reconstruct

the same excitation signal as is used in the encoder. The evolution (difference) between the previous speech segment and the currently encoded speech segment is further modelled using an innovation selected from a fixed codebook.

14. A problem of strong prediction inherent in CELP-based speech coders appears in presence of transmission errors (erased frames or packets) when the state of the encoder and the decoder become desynchronized. Due to the prediction, the effect of an erased frame is thus not limited to the erased frame, but continues to propagate after the erasure, often during several following frames.

15. Transitions from unvoiced speech segment to voiced speech segment (e.g. transition between a consonant or a period of inactive speech, and a vowel) or transitions between two different voiced segments (e.g. transitions between two vowels) are the most problematic cases for frame erasure concealment. When a transition from unvoiced speech segment to voiced speech segment (voiced onset) is lost, the frame right before the voiced onset frame is unvoiced or inactive and thus no meaningful periodic excitation is found in the buffer of the past excitation (adaptive codebook). At the encoder, the past periodic excitation builds up in the adaptive codebook during the onset frame and the following voiced frame is encoded using this past periodic excitation. Most frame error concealment techniques use the information from the last correctly received frame to conceal the missing frame. When the onset frame is lost, the decoder past excitation buffer will be thus updated using the noise-like excitation of the previous frame (unvoiced or inactive frame). The periodic part of the excitation is thus completely missing in the adaptive codebook at the decoder after a lost voiced onset and it can take up to several frames for the decoder to recover from this loss.

16. A similar situation occurs in the case of lost voiced to voiced transition. In that case, the excitation stored in the adaptive codebook before the transition frame has typically very different characteristics from the excitation stored in the adaptive codebook after the transition. Again, as the decoder usually conceals the lost frame with the use of the past frame information, the state of the encoder and the decoder will be very different, and the synthesized signal can suffer from distortion. These are issues that the patent seeks to address.

17. The patent includes four independent claims, claims 1, 10, 22 and 29 (with features labelled as in the observations of the patent proprietor – N.B. the labelled features included with the request do not correspond with the features as they appear in the granted claims), which read:

    1. *(1.1) A transition mode device for use in a predictive-type sound signal codec for producing a transition mode excitation replacing an adaptive codebook excitation*

        (1.1.1) in a transition frame and/or at least one frame following the transition in the sound signal, comprising:

        (1.2) an input for receiving a codebook index; and

        (1.3) a transition mode codebook for generating a set of codevectors independent from past excitation,

(1.3.1) the transition mode codebook being responsive to the codebook index for generating, in the transition frame and/or the at least one frame following the transition, one of the codevectors of the set corresponding to said transition mode excitation;

(1.3.2) wherein the transition mode codebook comprises a codebook of glottal impulse shapes.

10. (10.1) An encoder device for generating a transition mode excitation replacing an adaptive codebook excitation

(10.1.1) in a transition frame and/or at least one frame following the transition in a sound signal, comprising:

(10.2) a generator of a codebook search target signal;

(10.3) a transition mode codebook for generating a set of codevectors independent from past excitation,

(10.3.1) wherein the codevectors of said set each corresponds to a respective transition mode excitation and

(10.3.2) wherein the transition mode codebook comprises a codebook of glottal impulse shapes;

(10.4) a searcher of the transition mode codebook for finding the codevector of said set corresponding to a transition mode excitation optimally corresponding to the codebook search target signal.

22. (22.1) A transition mode method for use in a predictive-type sound signal codec for producing a transition mode excitation replacing an adaptive codebook excitation

(22.1.1) in a transition frame and/or at least one frame following the transition in the sound signal, comprising:

(22.2) providing a transition mode codebook for generating a set of codevectors independent from past excitation,

(22.2.1) wherein the transition mode codebook comprises a codebook of glottal impulse shapes;

(22.3) supplying a codebook index to the transition mode codebook; and

(22.4) generating, by means of the transition mode codebook and in response to the codebook index, one of the codevectors of the set corresponding to said transition mode excitation.

29. (29.1) An encoding method for generating a transition mode excitation replacing an adaptive codebook excitation

(29.1.1) in a transition frame and/or at least one frame following the transition in a sound signal, comprising:

(29.2) generating a codebook search target signal;

(29.3) providing a transition mode codebook for generating a set of codevectors independent from past excitation,

(29.3.1) the codevectors of said set each corresponding to a respective transition mode excitation,

(29.3.2) wherein the transition mode codebook comprises a codebook of glottal impulse shapes;

(29.4) searching the transition mode codebook for finding the codevector of said set corresponding to a transition mode excitation optimally corresponding to the codebook search target signal.

18. As indicated in the opening paragraph above, the request is directed to the validity of claims 1 and 10 alone. Therefore, my considerations will be limited to these two claims. However, it appears probable that any opinion concerning the validity of claims 1 and 10 would also be applicable to claims 22 and 29 as well. In fact, as recognised by the observer, the claim chart accompanying the request appears to confuse the features of claim 1 with those of claim 22. And, the features of the independent claims identified with the request appear to relate to features translated from the German claim wording of the granted EP patent, rather than the original English claim wording. However, it is possible to assess the features outlined above based on the comments of both the requester and observer.

Claim construction

19. Before I can determine an opinion as to the validity of the patent, I must first construe the claims. This means interpreting the claims in light of the description and drawings as instructed by section 125(1) of the Patents Act:

For the purposes of this Act an invention for a patent for which an application has been made or for which a patent has been granted shall, unless the context otherwise requires, be taken to be that specified in a claim of the specification of the application or patent, as the case may be, as interpreted by the description and any drawings contained in that specification, and the extent of the protection conferred by a patent or application for a patent shall be determined accordingly.

20. I must interpret the claims in context through the eyes of the person skilled in the art. Ultimately, the question is what the person skilled in the art would have understood the patentee to be using the language of the claims to mean. This approach has been confirmed in the recent decisions of the High Court in *Mylan v Yeda*[1] and the Court of Appeal in *Actavis v ICOS*[2].

21. Neither the requester nor the observer makes a submission regarding the identity of the person skilled in the art. However, I note that the preliminary opinion of the German Federal Patent Court considers the person skilled in the art to be an engineer in electrical engineering, information technology or communications engineering with a university degree (diploma or master's degree) and several years of professional experience as well as relevant knowledge in the field of digital signal processing, in particular the coding of audio data. I think that this paints a fair picture of the skilled person.

22. Based on the above, the general common knowledge of the skilled person would include the design and operation of typical CELP speech coder / decoder systems (N.B. paragraph 0042 of the patent acknowledges that "CELP is believed to be otherwise well known to those of ordinary skill in the art"). The concept of *transition* between speech segments would also be familiar to the skilled person. Hence, the issues identified in paragraphs 14 to 16 above would have been familiar to the skilled person.

23. Each of the independent claims includes reference to "*a codebook of glottal impulse shapes*". The requester has argued that, on the basis of the patent specification, the term "*glottal*" cannot be understood to have any clear technical characteristics intended to be imposed on the "*impulse shapes*". So, the requester submits that the term "*glottal impulse shapes*" must be understood as a functional feature, the function being to supplement or replace the adaptive codebook with a better codebook, which is thereby specifically applied to the (voiced onsets of a) transitional frame.

24. The observer contends that the patent explains in detail how such a "*codebook of glottal impulse shapes*" is constructed, e.g. at paragraph 0079:

> "*In principle, the glottal-shape codebook consists of quantized normalized shapes of the glottal impulses placed at a specific position*."

25. Furthermore, the observer states that the patent proposes to represent a glottal pulse as a function of 17 discrete values (see paragraph 0080) and to store in the "*codebook of glottal impulse shapes*" eight representative forms of such pulses (paragraph 0081). Specifically, a so-called unity impulse (or, understood to be a unit impulse) is not a '*glottal impulse shape*'; paragraph 0079 of the patent reads:

1 *Generics UK Ltd (t/a Mylan) v Yeda Research and Dev. Co. Ltd & Anor [2017] EWHC 2629 (Pat)*
2 *Actavis Group & Ors v ICOS & Eli Lilly & Co. [2017] EWCA Civ 1671*

"*In its simplest form, the shape of the glottal impulse can be represented by a unity impulse and does not need to be quantized. In that case, only its position in the subframe is determined. However the performance of such a simple codebook is very limited*."

26. Finally, the observer indicates that the patent requires a "*codebook of glottal impulse shapes*" (i.e. plural), which requires a codebook having more than one glottal impulse shape, rather than a unit impulse that represents only one shape. The purpose of the "*transition mode codebook*" is stated in paragraph 0023 of the patent:

    "*The idea behind the TM coding technique is thus to supplement the adaptive codebook with a better codebook populated with simplified quantized versions of glottal impulses to encode the voiced onsets."*

27. The requester responds by indicating that "*glottal*" is a term from biology, rather than a technical term used in audio coding. I note that the dictionary definition of "*glottal*" is relating to or produced by the glottis, which is the part of the larynx consisting of the vocal cords and the opening between them; it affects voice modulation through expansion and contraction.

28. It is my opinion that a skilled person would interpret the expression "*codebook of glottal impulse shapes*" in the light of the description to mean a codebook of pulses shaped to generate excitations that would synthesize in the coder the sounds produced by the human glottis in the larynx, particularly at the voice onset. Whilst I do not believe that the description rules out a unit impulse as one "*glottal impulse shape*", I believe that a skilled person would understand that the "*codebook of glottal impulse shapes*" (i.e. plural) would require several different "*impulse shapes*", rather than unit impulses alone.

Validity – novelty and inventive step

29. Section 1(1) of the Patents Act reads:

    A patent may be granted only for an invention in respect of the following conditions are satisfied, that is to say –

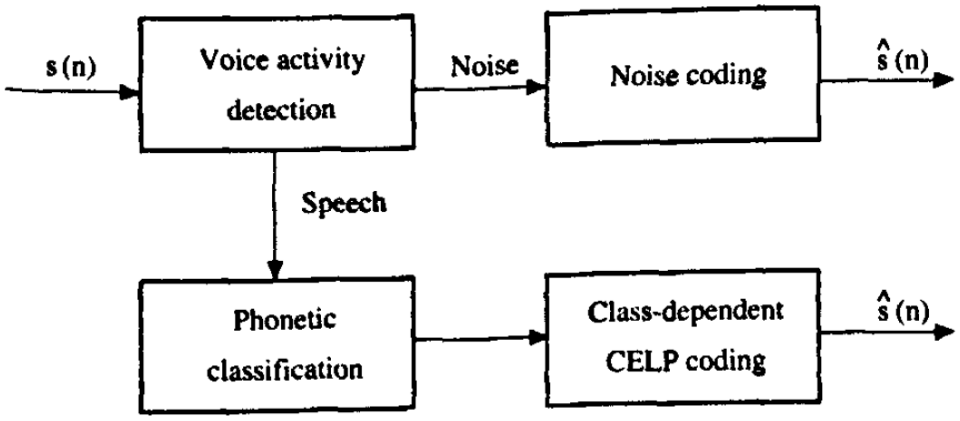
    (a) the invention is new;

    (b) it involves an inventive step…

30. The requester has argued that claims 1 and 10 of the patent are not novel over PAKSOY or DA SILVA. Alternatively, the requester argues that claims 1 and 10 of the patent lack an inventive step over PAKSOY in combination with LINDEN.

31. PAKSOY describes a variable bit-rate speech coder in which each frame of active speech is classified into voiced, unvoiced and onset categories. A simplified block diagram of the resulting coder, called *Variable Rate Phonetic Segmentation (VRPS)*, is illustrated in the figure reproduced below. If speech is detected, the input frame is phonetically classified into one of four possible categories and each category

corresponds to a CELP-type algorithm suited to the content of the speech frame.



"*Since VRPS is a CELP-type coder, the excitation parameters are shape vectors and gains selected from appropriate codebooks using an analysis-by-synthesis search technique. The properties of each shape codebook are determined by the properties of the phonetic class to be coded*."

33. Of particular significance, section 5.1.3 of PAKSOY discusses onset subframes. It explains:

"*during a transition from an unvoiced sound to a voiced sound, the glottal excitation undergoes a very rapid change… Consider the first voiced subframe detected following an unvoiced segment. During this subframe, the LPC residual (and hence the glottal excitation) contains the initial pitch pulses of the voiced segment which will generally continue for several subsequent frames. An adaptive codebook will be unsuccessful in capturing these pitch pulses since the previous subframe is unvoiced and the past excitation does not contain any pitch pulses. Therefore it will not offer a useful first approximation to the overall excitation… Furthermore, if the coding of the onset excitation fails, the adaptive codebook search performed in subsequent subframes will also have difficulty generating a good approximation of the periodic part of the LPC excitation. Consequently, poor coding of the onset will degrade the quality of the synthesized speech for several subframes.*"

34. The remedy provided by *VRPS*, as described in PAKSOY, is:

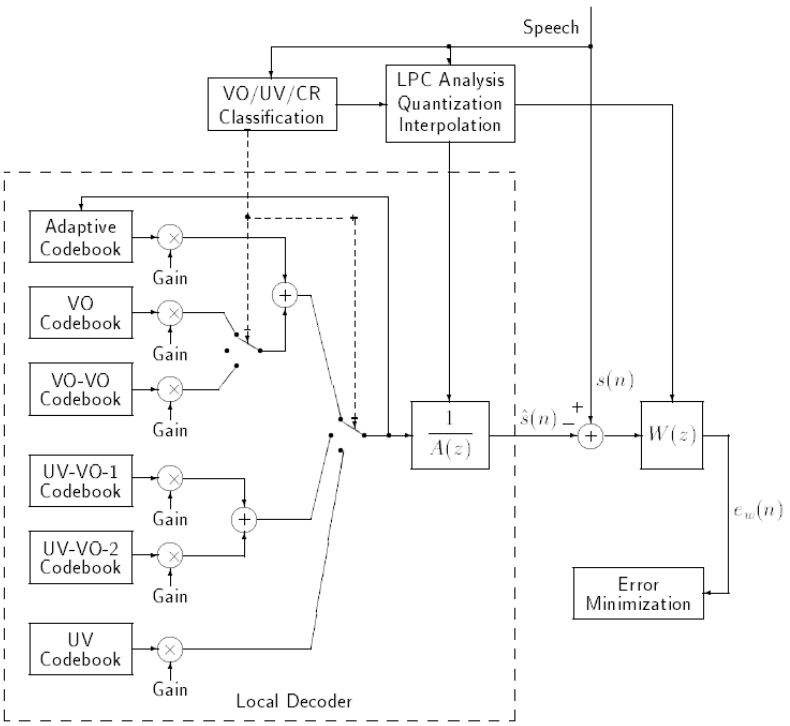
"*We code onsets using a two-stage excitation structure. Each stage is a 10 bit sparse codebook. The first stage is intended to accurately capture the pitch pulses. It contains ternary codewords with one and two non-zero pulses (+1 or -1), with a minimum pulse separation of 20 samples, as well as center-clipped, Gaussian vectors. The second stage must code the stochastic part of the excitation, and consists of center-clipped unit-variance, Gaussian vectors. The gain values corresponding to the first and second onset codebooks are scalar quantized at 5 and 4 bits respectively*."

35. The observer has argued that the "first stage" codebook of PAKSOY comprises code

vectors with unit impulses at different positions and so does not constitute a "*codebook of glottal impulse shapes*". Furthermore, the observer states that the "center-clipped, Gaussian vectors" mentioned in connection with the first stage codebook do no comprise "*glottal impulse shapes*" either, arguing that the resulting vectors represent noise-like signal segments.

36. However, I am in agreement with the preliminary opinion of the German Federal Patent Court in coming to the view that PAKSOY discloses each of the features of claim 10 of the patent. I am of the opinion that the "first stage" codebook described in PAKSOY does meet the requirements of the "*codebook of glottal impulse shapes*" as required by the independent claims of the patent. PAKSOY describes this "first stage" codebook as 'accurately capturing the pitch pulses' in contrast to a conventional adaptive codebook, which "will be unsuccessful in capturing these pitch pulses since the previous subframe is unvoiced and the past excitation does not contain any pitch pulses." In particular, it appears clear to me that a skilled person would understand the purpose of the "first stage" codebook provided by *VRPS* in PAKSOY is to represent "the glottal excitation… the initial pitch pulses of the voiced segment". Hence, I am of the opinion that PAKSOY discloses each and every feature of claims 1 and 10 of the patent.

37. The German Federal Patent Court came to the preliminary opinion that claims 1 and 10 of the patent are new over DA SILVA. I do not intend to revisit all of the considerations of the Federal Patent Court (particularly where these are not disputed by the observer) but, in summary, whilst DA SILVA does describe (as illustrated in the figure reproduced below) a CELP-based scheme that employs different combinations of codebooks for different categories of "sub-blocks" (i.e. subframes) of speech segments (e.g. "UV-VO" category for "unvoiced" to "voiced" transition, "VO-VO" category for "voiced" to "voiced" transition, etc.), once again the argument raised by the observer is that none of the codebooks of DA SILVA comprises "*glottal impulse shapes*".



38. The requester points to section 3.3 of DA SILVA in which, for "unvoiced-voiced type" segments:

> "*the adaptive codebook is not activated. These segments employ two codebooks: UV-VO-1 containing 256 vectors and UV-VO-2 with 128 vectors. The first one is usually responsible for the large major pulses in the LPC excitation, which are of major importance for the quality of reconstructed speech.*"

39. The requester equates the "UV-VO-1" codebook of DA SILVA to the "*codebook of glottal impulse shapes*" of the patent. In particular, they argue that the "*large major pulses in the LPC excitation*" are equivalent to the glottal impulses of the patent (with reference to par.0075).

40. Although the disclosure of DA SILVA is not as clear as that in PAKSOY (at least as far as "*glottal impulse shapes*" is concerned), I do find some indication in DA SILVA that the "UV-VO-1" codebook would inevitably lead to such impulse shapes. In section 3.2, DA SILVA describes the optimization of the five fixed codebooks used for excitation modelling, including the "UV-VO-1" codebook. This section indicates that the 48-element "codevectors" in each codebook have 8 nonzero elements and that, whilst each codebook "initially contains arbitrary sparse codevectors":

> "*Then, it is iteratively optimized to minimize the perceptually weighted distortion between the original and the CELP-synthesized speech signals. This optimization is done using an appropriate speech training data base.*"

41. Therefore, unlike the German Federal Patent Court, in my opinion this optimization of the codebooks will inevitably lead to "*glottal impulse shapes*" representing the different categories of transitions (unvoiced-to-voiced, voiced-to-voiced) – I also note the similar training of the "*glottal-shape codebook*" as outlined in paragraph 0081 of the patent – such that DA SILVA discloses each feature of claims 1 and 10 of the patent.

42. Since I have opined that claims 1 and 10 are not novel over PAKSOY, it is not necessary for me to consider inventive step in view of PAKSOY in combination with LINDEN.

Opinion

43. It is my opinion that European Patent (UK) 2102619 B1 is invalid as claims 1 and 10 are not new in the light of either PAKSOY or DA SILVA. Although not subject of this request, it seems likely that claims 22 and 29 are also not new.

Application for review

44. Under section 74B and rule 98, the proprietor may, within three months of the date of issue of this opinion, apply to the comptroller for a review of the opinion.

Dan Hickery
Examiner

NOTE

This opinion is not based on the outcome of fully litigated proceedings. Rather, it is based on whatever material the persons requesting the opinion and filing observations have chosen to put before the Office.