

**Internal distribution code:**

- (A)  Publication in OJ  
(B)  To Chairmen and Members  
(C)  To Chairmen  
(D)  No distribution

**Datasheet for the decision  
of 1 February 2010**

**Case Number:** W 0001/09 - 3.5.04

**Application Number:** 07834923.0

**Publication Number:** -

**IPC:** H04N 7/15

**Language of the proceedings:** EN

**Title of invention:**

System and procedure of hands free speech communication using  
a microphone array

**Applicant:**

Micronas Nit

**Opponent:**

-

**Headword:**

-

**Relevant legal provisions:**

PCT Art. 17(3)(a)  
PCT R. 13, 40.1, 40.2(c),(e)

**Relevant legal provisions (EPC 1973):**

-

**Keyword:**

"Lack of unity a priori"

**Decisions cited:**

W 0016/08

**Catchword:**

-



**Case Number:** W 0001/09 - 3.5.04

**International Application No.** PCT/RS2007/000017

**D E C I S I O N**  
**of the Technical Board of Appeal 3.5.04**  
**of 1 February 2010**

**Appellant:** Micronas Nit  
Fruskogorska 11a  
RS 21000 Novi Sad (RS)

**Representative:** -

**Decision under appeal:** Protest according to Rule 40.2(c) of the Patent Cooperation Treaty made by the applicants against the invitation (payment of additional fees) of the European Patent Office (International Searching Authority) dated 30 September 2008.

**Composition of the Board:**

**Chairman:** F. Edlinger  
**Members:** C. Kunzelmann  
T. Bokor

## Summary of Facts and Submissions

- I. The international application PCT/RS2007/000017 was filed with 24 claims. The independent claims 1, 5 and 12 read as follows.

Claim 1:

"The system for a hands-free speech communication using a microphone array, which contains a digital TV receiver that allows audio and video communication facilities in full duplex wherein the digital TV receiver (100) performs a stereo audio reproduction (102) of the stereo TV signal and a mono reproduction of an incoming speech signal needed for a video-telephone communication, and which has a moving video camera (104) for a speaker's recording in a room and presenting a picture of the remote speaker on a window of its screen (105); which contains a microphone system embedded (103) in the TV receiver (100) that records the voice of the speaker and other surrounding sounds at the near end, and that has the purpose to locate the position of the speaker in the room and to control the direction of the video camera (104)."

Claim 5:

"The systems wherein it cancels acoustic echo (209) that is generated in stereo loudspeakers (102) of the TV set and is composed of both a stereo audio TV signal (205) and a mono speech signal that originates from a far-end speaker (204)."

Claim 12:

"The technique for hands-free full duplex speech communication using microphone arrays, wherein it performs parallel processing of microphone signals generated in the microphone array and thus adaptively cancels acoustic echo in the microphone signals, performs the direction of arrival estimation of the direct sound wave of the near-end speaker, forms a superdirective beamforming characteristic of the microphone array and controls its azimuth coordinate, suppresses all noise signals contained in the microphone signals and performs an automatic control of the level of the transmitted voice signal."

- II. With an "invitation to pay additional fees" (Form PCT/ISA/206) dated 30 September 2008 the European Patent Office (EPO), acting in its capacity as International Searching Authority (ISA) under Article 16 PCT and Article 152 EPC, informed the applicant that it had found two groups of inventions in the international application and that it considered that the international application did not comply with the requirement of unity of invention (Rules 13.1, 13.2 and 13.3 PCT). The applicant was invited to pay a fee for one additional invention in accordance with Article 17(3) (a) PCT and Rule 40.1 PCT.
  
- III. On 30 October 2008 the applicant paid the additional fee under protest accompanied by a reasoned statement to the effect that the international application complied with the requirement of unity of invention.

- IV. On 19 December 2008 the ISA mailed an "invitation to pay a protest fee" (Form PCT/ISA/228), since a prior review of the justification for the "invitation to pay additional fees" had resulted in the requirement of payment of one additional fee being upheld because the "invitation to pay additional fees" was justified. The result of the prior review was annexed to the "invitation to pay a protest fee".
- V. The protest fee was paid on 19 January 2009.
- VI. The reasons given in the "invitation to pay additional fees" as to why the international application was considered not to comply with the requirement of unity of invention can be summarised as follows.

The international application claimed a first group of inventions (group I, claims 1 to 11) and a second group of inventions (group II, claims 12 to 24).

Group I related to a hands-free full duplex video conference system that made use of a digital TV receiver provided with stereo audio means that allowed the visualization of a picture of a remote speaker. The system was provided with an integrated array of microphones having the purpose of locating the position of the local speaker with the aim of controlling the direction of an integrated video camera.

Group II related to a hands-free full duplex speech communication system provided with a means that processed in parallel the acoustic signal provided by an array of microphones in order to estimate the direction of arrival of the local user's voice, to

cancel acoustic echoes and environmental noises and to automatically control the level of the transmitted voice signal.

For the following reasons it was considered that the international application did not comply with the requirement of unity of invention.

The common features of the respective independent claims of both groups were a hands-free full duplex communications system, an array of microphones and acoustic signals provided by the array of microphones being processed in order to estimate the direction of arrival of the local user's voice.

These features were *a priori* well-known. Examples of prior art disclosing these features were US2006/0132595, US6593956 and WO03043327. Thus these features were not special technical features within the meaning of Rule 13.2 PCT.

For group I the special technical features were the use of a digital TV receiver provided with stereo audio means that allowed the visualization of a picture of a remote speaker. Furthermore the estimated position of the local speaker was used for controlling the direction of an integrated video camera.

For group II the special technical features were the array of microphones which formed a controllable superdirective beamforming characteristic advantageous for the cancellation of acoustic echoes and environmental noise and an automatic control of the level of the transmitted voice signal.

The special technical features of groups I and II were not the same or corresponding as they had different effects. In group I the objective problem was the capturing and displaying of images of the remote speaker. In group II the objective problem was the enhancing of the quality of the transmitted audio signal.

VII. The applicant's protest can be summarised as follows.

The invention could not be divided in two independent groups. Both claim groups I (claims 1 - 11) and II (claims 12 - 24) included both audio signal processing for speech enhancement as well as speaker localisation needed for video camera pointing. The speaker localisation was closely related to the speech processing algorithms. Namely echo interference cancellation was crucial for correct speaker localisation when a TV audio signal was present, and the super-directive beamformer needed the correct position of the speaker in order to enhance the speech signal and to suppress ambient noise.

The common special technical features of claims groups I and II were echo cancellation used to improve speaker localization in presence of a strong TV signal echo and super-directive beam forming based on the estimated speaker localisation. The invention was designed for domestic use and allowed a first conversation participant watching a TV program without reducing the TV sound volume and at the same time allowed the first conversation participant to speak to a second conversation participant at a distance. All

the process modules of the invention were customised so that the conversation participants were not disturbed by the TV sound volume on the other participant's side. Most of the prior art documents found in the partial international search instead concerned systems for use under the controlled working conditions of a video conference in which only one person was speaking at a time.

## **Reasons for the Decision**

### *1. Competence and admissibility*

1.1 The application in suit was filed on 19 September 2007. Therefore the protest is subject to the provisions of the PCT as in force from 1 April 2007. The Boards of Appeal are responsible for deciding on protests relating to PCT applications pending at the time of entry into force of the EPC 2000 (13 December 2007), see Article 1(6) of the Decision of the Administrative Council of 28 June 2001 on the transitional provisions under Article 7 of the Act revising the European Patent Convention of 29 November 2000. Details of the procedure are guided by the Decision of the President of the EPO dated 24 June 2007, Article 3 (OJ EPO 2007, Special edition No. 3, 140). For more details see also W 16/08, points 1.1-1.5 of the reasons.

1.2 The protest fee has been paid in time, and the protest contains a reasoned statement as to why the inventions for which the additional search fees have been paid fulfil the requirement of unity. Accordingly, the



protest was properly made and it is admissible (Rule 40.2 (c) and (e) PCT).

2. *The technical field of the application*

2.1 The international application concerns the technical field of acoustic signal processing, in particular acoustic echo cancellation and the location and selection of an active speaker in the presence of noise and reverberations in the acoustic environment (see page 1, lines 5 to 9). Such acoustic signal processing is used in hands-free full-duplex speech communication systems such as video-phone systems, teleconference systems, hands-free systems for use in cars, etc. (see page 1, lines 11 to 14). In particular, the location and selection of an active speaker, that is the separation of the desired acoustic signal from disturbances, may conventionally be solved by using a microphone array having a number of microphones arranged in line a small distance from each other. With appropriate processing of the signals of such a conventional microphone array, a direction dependent sensitivity of the microphone system is achieved (see page 1, line 33 to page 2, line 6). In video-phone and teleconferencing systems in particular it may be important to determine the direction of the speaker to the microphone array in order to control a moveable camera so that it points towards the speaker (see page 2, lines 7 to 15).

2.2 The application acknowledges that hands-free full-duplex speech communication systems of the above type belong to the background art and lists a number of documents dealing with the problems of noise reduction,

echo cancellation, adaptive beamforming, talker localisation, etc., either alone or in specific combinations (see page 3, lines 12 to 30).

3. *The claims*

3.1 Independent claim 1 defines a system containing a digital TV receiver that allows audio and video communication facilities and performs a stereo audio reproduction of the stereo TV signal and a mono reproduction of an incoming speech signal needed for a video-telephone communication. In the system of claim 1 some acoustic signal processing of the kind described in point 2.1 above may be performed, in particular for the control of the direction of a video camera used for the video-telephone communication.

3.2 Independent claim 5 specifies a more general system for performing some acoustic signal processing of the kind described in point 2.1 above. Also the system of claim 5 comprises a TV set which may perform both a stereo audio TV signal and a mono speech signal that originates from a far-end speaker. Claim 5 was considered to relate to the same group of inventions as claim 1 in the invitation to pay additional fees and thus need not be dealt with separately in the present decision.

3.3 Independent claim 12 defines a technique for hands-free full duplex speech communication using microphone arrays. The technique adaptively cancels acoustic echoes in the microphone signals, performs the direction of arrival estimation of the direct sound

wave of the near end-speaker and other acoustic signal processing.

4. *The groups of inventions in the claims*

4.1 When read in the context of the application as a whole (see section 2 above), claims 1 and 5 relate to acoustic processing performed in the specific context of using a (digital) stereo TV system to carry out speech (and video) communication. The technical problem addressed by these features concerns an improvement of determining the location, in the complex situation of these different audio and accompanying noise sources, of a near-end speaker for recording the voice and moving a camera in the speaker's direction (see page 4, lines 7 to 16, and page 7, lines 3 to 17). These independent claims thus fall within "group I" identified in the invitation to pay additional fees.

4.2 Independent claim 12 however does not refer to the same specific context as claims 1 and 5. Instead claim 12 generally specifies acoustic signal processing on a functional level. None of the functions specified in claim 12 is specific to the use of a stereo TV system for carrying out speech (and video) communication. Instead all of the functions specified in claim 12 may also occur, for instance, in the context of video-phone or teleconference systems. The features of claim 12 address the technical problem of improving the directivity of the microphone array and the quality of the audio signal transmitted in a full-duplex speech communication (by adaptively cancelling acoustic echoes and automatic control of the level of the transmitted voice signal; see page 1, lines 22 to 28; page 2,

lines 23 to 27; page 3, lines 6 to 9; page 5, lines 24 to 29; page 11, lines 32 to 37). This claim thus falls within "group II" identified in the invitation to pay additional fees.

- 4.3 Hence at least claims 1 and 5 on the one hand and claim 12 on the other hand relate to different inventions. The potential special technical features of these inventions in view of the background art indicated in the description (hands-free full-duplex speech communication and the related problems; see point 2.2 above) address different technical problems, as stated in the invitation to pay additional fees. Although some of the dependent claims and the description show that both problems may arise together in the stereo audio TV signal reproduction situation to which claim 1 relates, and that the features of the different solutions may be combined, the inventions as set out in the independent claims are not so linked as to form a single general inventive concept. Therefore the invitation to pay an additional fee was correct. Furthermore the lack of unity was directly evident *a priori*, that is before considering the claims in relation to any prior art. In the invitation to pay additional fees prior art documents were merely cited as examples confirming the ISA's assessment that the international application did not comply with the requirement of unity of invention in view of the lack of the same or corresponding special technical features of the two groups of inventions.

5. *The protest*

5.1 The protest is mainly based on the argument that both claim groups included both audio signal processing for speech enhancement as well as speaker localisation needed for video camera pointing. This argument however does not take into account that the claims do not specify corresponding special technical features, but individually relate to different aspects of problems which may, but not necessarily do, arise in common. Specifically, claim 1 does not set out features relating to a particular processing for speech enhancement. Instead claim 1 focuses on the aspect of speaker localisation in the presence of a stereo audio reproduction of the stereo TV signal and a mono reproduction of an incoming speech signal. Claim 12 specifies both speech enhancement, for instance acoustic echo cancellation, and speaker localisation. But in claim 12 speaker localisation is not necessary for video camera pointing, since claim 12 neither mentions a camera nor mentions video communication. Nor does it relate to a situation where stereo audio reproduction of the stereo TV signal causes a particular problem for transmitting a voice signal.

Also the description makes clear that adaptive acoustic echo cancellation, an adaptive directivity characteristic of the microphone array, speaker localisation and automatic gain control are different specific aspects of the invention(s). The improvement of each of these specific aspects is a "specialty" of the invention(s) (see page 4, line 20, to page 5, line 36). Even though the description also indicates that the invention in essence "is one optimally

designed algorithm" and that an improvement consists in the "integration process of all algorithms", the application as a whole discloses that the different specific aspects may be considered separately.

5.2 Also the applicant's argument that echo cancellation was used to improve speaker localization in the presence of a strong TV signal echo and super-directive beam forming was based on the estimated speaker localisation is not based on special technical features of the claims.

5.3 The applicant's argument that the invention was designed for domestic use is not based on the actual wording of the claims, either. Furthermore the description (see page 15, lines 10 to 13) also specifies that the invention "relies on free speech communication in one digital television system, but at the same time it can be used for others (*sic*) communication systems, as are video-phone systems, teleconference systems, speakerphones in the room or car, human-computer voice communication, etc.". Thus the application as a whole also makes clear that it is not limited to one invention or one group of inventions in which all process modules are customised so that conversation participants are not disturbed by the TV sound volume on the other participant's side.

5.4 Hence the applicant's protest is not justified.

**Order**

**For these reasons it is decided that:**

The protest is dismissed.

The Registrar:

The Chairman:

L. Fernández Gómez

F. Edlinger