

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

**AMAZON.COM, INC.,
AMAZON.COM SERVICES LLC, and
AMAZON WEB SERVICES, INC.,**
Petitioners,

v.

SOUNDCLEAR TECHNOLOGIES LLC,
Patent Owner.

Case No. IPR2025-01096
Patent No. 9,031,259

DECLARATION OF RICHARD STERN, PH.D.

<i>Amazon v. SoundClear</i> US Patent 9,031,259 Amazon EX-1002
--

TABLE OF CONTENTS

I.	BACKGROUND -----	1
A.	Experience and Qualifications-----	1
B.	Materials Considered -----	3
II.	APPLICABLE LEGAL STANDARDS -----	4
A.	Claim Construction -----	5
B.	Obviousness -----	6
III.	PERSON OF ORDINARY SKILL IN THE ART -----	10
IV.	TECHNOLOGY BACKGROUND -----	12
A.	Detecting Speech in Audio Signals Was Known -----	12
B.	Determining an Audio Signal's Incoming Direction Was Known -----	13
C.	Reducing Noise via Adaptive Filters Was Known -----	13
D.	Using Voice Activity Detection, DOA Determination, and an Adaptive Filter in the Same Device Was Known -----	15
V.	THE '259 PATENT -----	17
A.	Overview -----	17
B.	Claims -----	19
VI.	CLAIMS 1-20 OF THE '259 PATENT WOULD HAVE BEEN OBVIOUS -----	20
A.	Claims 1-3, 5-6, 8-9, 12-15, and 17-20 Would Have Been Obvious in View of Chen and Li-----	20
1.	Claim 1 -----	23
a.	Preamble -----	23

b.	Element 1[a]: Speech Segment Determiner -----	24
c.	Element 1[b]: Voice Direction Detector -----	26
d.	Element 1[c]: Adaptive Filter -----	27
e.	Motivation to Combine -----	30
2.	Claim 13 -----	31
3.	Claims 2 and 14-----	33
4.	Claims 3 and 15-----	37
5.	Claims 5 and 17-----	40
6.	Claims 6 and 18-----	41
7.	Claims 8 and 19-----	43
8.	Claims 9 and 20-----	45
9.	Claim 12 -----	46
B.	Claims 1-3, 5-6, 8-9, 12-15, and 17-20 Would Have Been Obvious in View of Chen, Li, and Florencio-----	49
1.	Claim 1 -----	51
2.	Claim 13 -----	55
3.	Claims 2 and 14-----	55
4.	Claims 5 and 17-----	56
5.	Claims 3, 6, 8-9, 12, 15, and 18-20 -----	57
C.	Claims 4, 7-8, 12, 16, and 19-20 Would Have Been Obvious in View of (i) Chen, Li, and Visser or (ii) Chen, Li, Florencio, and Visser -----	57
1.	Claims 4 and 16-----	58

2.	Claim 7 -----	61
3.	Claims 8 and 19-----	63
4.	Claim 12 -----	67
5.	Claim 20 -----	69
D.	Claims 10 and 11 Would Have Been Obvious in View of (i) Chen, Li, Kim, and Kleffner or (ii) Chen, Li, Florencio, Kim, and Kleffner-----	69
1.	Claim 10 -----	69
2.	Claim 11 -----	73
E.	Claims 1-3, 9, and 13-15 Would Have Been Obvious in View of Strömmer -----	74
1.	Claim 1 -----	76
a.	Preamble -----	76
b.	Element 1[a]: Speech Segment Determiner -----	76
c.	Element 1[b]: Voice Direction Detector -----	79
d.	Element 1[c]: Adaptive Filter -----	82
i.	Beamformer-----	82
ii.	Noise Reduction-----	83
2.	Claim 13 -----	85
3.	Claims 2 and 14-----	85
4.	Claims 3 and 15-----	87
5.	Claim 9 -----	90
F.	Claims 4, 12, and 16 Would Have Been Obvious in View of Strömmer and Visser -----	91

1.	Claims 4 and 16-----	92
2.	Claim 12 -----	92
G.	Claims 5-6 and 17-18 Would Have Been Obvious in View of Strömmer and Brandstein-----	93
1.	Claims 5 and 17-----	94
2.	Claims 6 and 18-----	98
H.	Claims 7-8 and 19-20 Would Have Been Obvious in View of Strömmer, Brandstein, and Visser-----	100
1.	Claim 7 -----	100
2.	Claims 8 and 19-----	101
3.	Claim 20 -----	102
I.	Claims 10 and 11 Would Have Been Obvious in View of Strömmer, Kim, and Kleffner -----	103
1.	Claim 10-----	103
2.	Claim 11 -----	104
J.	Claims 1-3, 9, and 13-15 Would Have Been Obvious in View of Florencio-----	105
1.	Claim 1 -----	107
a.	Preamble -----	107
b.	Element 1[a]: Speech Segment Determiner -----	108
c.	Element 1[b]: Voice Direction Detector-----	109
d.	Element 1[c]: Adaptive Filter -----	111
2.	Claim 13 -----	113
3.	Claims 2 and 14-----	113

4.	Claims 3 and 15-----	114
5.	Claim 9 -----	115
K.	Claims 4, 12, and 16 Would Have Been Obvious in View of Florencio and Visser-----	116
1.	Claims 4 and 16-----	117
2.	Claim 12 -----	118
L.	Claims 5-6 and 17-18 Would Have Been Obvious in View of Florencio and Brandstein -----	118
1.	Claims 5 and 17-----	119
2.	Claims 6 and 18-----	119
M.	Claims 7-8 and 19-20 Would Have Been Obvious in View of Florencio, Brandstein, and Visser -----	121
1.	Claim 7 -----	121
2.	Claims 8 and 19-----	122
3.	Claim 20 -----	123
N.	Claims 10 and 11 Would Have Been Obvious in View of Florencio, Kim, and Kleffner-----	123
VII.	SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS -----	125
VIII.	CONCLUSION -----	125

I, Richard Stern, Ph.D., do hereby declare:

1. I am making this declaration at the request of Petitioners Amazon.com, Inc., Amazon.com Services LLC, and Amazon Web Services, Inc. (collectively, “Amazon”). I have been retained by Amazon as a technical expert in this matter.

2. I am being compensated for my work on this case. My compensation does not depend on the content of this Declaration or the outcome of these proceedings. As part of a diversified portfolio, I own shares of Amazon stock. A financial professional manages my portfolio, and I do not. This stock ownership constitutes less than 1 percent of the total portfolio, and does not in any way affect the contents of my opinions in this Declaration.

I. BACKGROUND

A. Experience and Qualifications

3. I am a Professor at Carnegie Mellon University in the Department of Electrical and Computer Engineering, the Department of Computer Science, and the Language Technologies Institute. I have been on the faculty of Carnegie Mellon since 1977.

4. I received the S.B. degree from the Massachusetts Institute of Technology (MIT) in 1970, the M.S. from the University of California, Berkeley, in 1972, and the Ph.D. from MIT in 1977, all in electrical engineering.

5. While I was a graduate student, I worked as a Teaching and Research Assistant in the Department of Electrical Engineering at MIT, from 1973 to 1976. My teaching experience was in the areas of signal processing and acoustics. My research at MIT had been in the area of auditory processing and perception.

6. I am a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Acoustical Society of America, and the International Speech Communication Association (ISCA). I was the ISCA 2008-2009 Distinguished Lecturer, a recipient of the Allen Newell Award for Research Excellence in 1992, and I served as the General Chair of Interspeech 2006. Interspeech is the world's largest technical conference focused on speech processing and application.

7. Much of my current research is in spoken language systems, where I am particularly concerned with the development of techniques with which automatic speech recognition can be made more robust with respect to changes in environment and acoustical ambience. Furthermore, as a Professor, I have taught 15 courses, many of which are at the graduate level, in the general areas of signal processing, communication theory, auditory perception, and acoustics.

8. My relevant publications are available on Carnegie Mellon's web site at <http://www.cs.cmu.edu/afs/cs/user/robust/www/papers.html>.

9. I understand a copy of my current curriculum vitae, which lists my publications for the last ten years, is being submitted as Exhibit 1008.

B. Materials Considered

10. In preparing this Declaration, I have considered the following materials:

Exhibit No.	Description
1001	U.S. Patent No. 9,031,259
1003	U.S. Patent App. Publ. No. 2010/0081487 (“Chen”)
1004	U.S. Patent App. Publ. No. 2012/0197638 (“Li”)
1005	U.S. Patent App. Publ. No. 2008/0240463 (“Florencio”)
1006	Lae-Hoon Kim, Statistical Model Based Multi-Microphone Speech Processing: Toward Overcoming Mismatch Problem (2010) (Dissertation, University of Illinois at Urbana-Champaign) (“Kim”)
1007	PCT Patent App. Publ. No. WO2009/151578 (“Kleffner”)
1009	U.S. Patent App. Publ. No. 2009/0164212 (“Chan”)
1010	U.S. Patent App. Publ. No. 2013/0013303 (“Strömmer”)
1011	Excerpts from MICROPHONE ARRAYS: SIGNAL PROCESSING TECHNIQUES AND APPLICATIONS (Michael Brandstein et al. eds., 2001) (“Brandstein”)
1012	U.S. Patent No. 8,897,455 (“Visser”)
1013	Eur. Telecomm. Standards Inst. [ETSI], <i>Digital Cellular Telecommunications System (Phase 2+); Voice Activity Detector (VAD) for Adaptive Multi-Rate (AMR) Speech Traffic Channels; General Description (GSM 06.94 Version 7.1.1 Release 1998)</i> , ETSI EN 301 708 V7.1.1 (1999-12) (“ETSI”)
1014	U.S. Patent No. 7,359,504 (“Reuss”)
1015	U.S. Patent App. Publ. No. 2011/0264447 (“Visser ’447”)

Exhibit No.	Description
1016	Excerpts from SPRINGER HANDBOOK OF SPEECH PROCESSING (Jacob Benesty et al. eds., 2008) (“Benesty”)
1017	Excerpts from DAG STRANNEBY & WILLIAM WALKER, DIGITAL SIGNAL PROCESSING AND APPLICATIONS (2d ed. 2004) (“Stranneby”)
1018	Excerpts from COMPREHENSIVE DICTIONARY OF ELECTRICAL ENGINEERING (Phillip A. Laplante ed., 2d ed. 2005) (“Comprehensive Dictionary”)
1019	Bernard Widrow et al., <i>Adaptive Noise Cancelling: Principles and Applications</i> , 63 Proc. IEEE 1692 (1975) (“Widrow”)
1020	File History of U.S. Patent No. 9,031,259

11. I have reviewed the full file history of the ’259 patent. I have also relied on my education, training, and experience, and my knowledge of pertinent literature in the field of the ’259 patent.

II. APPLICABLE LEGAL STANDARDS

12. I have been asked to provide my opinion as to whether the claims of the ’259 patent would have been obvious to a person of ordinary skill in the art at the time of the alleged invention, in view of the prior art.

13. I am an electrical engineer by training and profession. The opinions I am expressing in this report involve the application of my training and technical knowledge and experience to the evaluation of certain prior art with respect to the ’259 patent.

14. Although I have been involved as a technical expert in patent matters before, I am not an expert in patent law. Therefore, the attorneys from Knobbe, Martens, Olson & Bear, LLP have provided me with guidance as to the applicable patent law in this matter. The paragraphs below express my understanding of how I must apply current principles related to patent validity to my analysis.

A. Claim Construction

15. It is my understanding that in determining whether a patent claim is obvious in view of the prior art, the Patent Office construes the claim by giving the claim terms their plain and ordinary meaning, as they would have been understood by a person of ordinary skill in the art at the time of the invention in view of the intrinsic record (patent specification and file history). For the purposes of this review, and to the extent necessary, I have interpreted each claim term in accordance with its plain and ordinary meaning as it would have been understood by a person of ordinary skill in the art at the time of the invention, in view of the intrinsic record. I have been instructed to assume that the time of the invention is September 15, 2011, which I understand to be the earliest claimed priority date of the '259 patent.

16. I understand that a patent and its prosecution history are considered “intrinsic evidence” and are the most important sources for interpreting claim language in a patent. I also understand that in reading the claim, I must not import limitations from the specification into the claim terms; in other words, I must not narrow the

scope of the claim terms by implicitly adding disclosed limitations that have no express basis in the claims. The prosecution history of related patents and applications can also be relevant.

17. I understand that sources extrinsic to a patent and its prosecution history (such as dictionary definitions and technical publications) may also be used to help interpret the claim language, but that such extrinsic sources cannot be used to contradict the unambiguous meaning of the claim language that is evident from the intrinsic evidence.

18. Unless expressly stated herein, I have applied the plain and ordinary meaning of the claim terms, which I understand is the meaning that a person of ordinary skill in the art would have given to terms in September 2011 based on a review of the intrinsic evidence.

B. Obviousness

19. It is my understanding that a claim is “obvious” if the claimed subject matter as a whole would have been obvious to a person of ordinary skill in the art at the time of the alleged invention. I understand that an obviousness analysis involves a number of considerations. I understand that the following factors must be evaluated to determine whether a claim would have been obvious: (i) the scope and content of the prior art; (ii) the differences, if any, between each claim of the ’259 patent and the prior art; (iii) the level of ordinary skill in the art in September 2011; and (iv)

additional considerations, if any, that indicate that the invention was obvious or not obvious. I understand that these “additional considerations” are often referred to as “secondary considerations” of non-obviousness or obviousness.

20. I also understand that the frame of reference when evaluating obviousness is what a hypothetical person of ordinary skill in the pertinent art would have known in September 2011. I understand that the hypothetical person of ordinary skill is presumed to have knowledge of all pertinent prior art references.

21. It is my understanding that something is “inherent in,” and therefore taught by, the prior art, if it necessarily flows from the explicit disclosure of the prior art. I understand that the fact that a certain result or characteristic *may be* present in the prior art is not sufficient to establish inherency. However, if the result or characteristic is *necessarily* present based upon the explicit disclosure in the prior art, it is inherent in the prior art and is therefore disclosed.

22. I understand that a prior art reference may be a pertinent prior art reference (or “analogous art”) if it is in the same field of endeavor as the patent or if it is pertinent to the problem that the inventors were trying to solve. A reference is reasonably pertinent if it logically would have commended itself to an inventor’s attention in considering the problem at hand. If a reference relates to the same problem as the claimed invention, that supports use of the reference as prior art in an obviousness analysis. Here, all of the references relied on in my obviousness analysis

below are in the same field of endeavor as the '259 patent, e.g., sound signal processing and/or automatic speech recognition. The references are also pertinent to a particular problem the inventor was focused on, e.g., reducing noise in audio signals containing speech.

23. It is my understanding that the law recognizes several rationales for combining references or modifying a reference to show obviousness of claimed subject matter. Some of these rationales include:

- combining prior art elements according to known methods to yield predictable results;
- simple substitution of one known element for another to obtain predictable results;
- a predictable use of prior art elements according to their established functions;
- using known techniques to improve similar devices (methods, or products) in the same way;
- applying a known technique to a known device (method, or product) ready for improvement to yield predictable results;
- choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success (in which case a claim would have been obvious to try);

- known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations would have been predictable to one of ordinary skill in the art; and
- some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

24. I understand that “secondary considerations” must be considered as part of the obviousness analysis when present. I further understand that the secondary considerations may include: (1) a long-felt but unmet need in the prior art that was satisfied by the claimed invention; (2) the failure of others; (3) skepticism by experts; (4) commercial success of a product covered by the patent; (5) unexpected results achieved by the claimed invention; (6) industry praise of the claimed invention; (7) deliberate copying of the invention; and (8) teaching away by others. I also understand that evidence of the independent and nearly simultaneous “invention” of the claimed subject matter by others is a secondary consideration supporting an obviousness determination and may support a conclusion that a claimed invention was within the knowledge of a person of ordinary skill as of September 2011. I am not aware of any evidence of secondary considerations that would suggest that the claims of the ’259 patent would have been nonobvious in September 2011.

25. I understand that when assessing obviousness, using hindsight is impermissible; that is, what is known today or what was learned from the teachings of the patent should not be considered. The patent should not be used as a road map for selecting and combining items of prior art. Rather, obviousness must be considered from the perspective of a person of ordinary skill at the time the alleged invention was made – September 2011 in this case.

26. I also understand that an obviousness analysis must consider the invention as a whole, as opposed to just a part or element of the invention. I understand this “as a whole” assessment to require showing that one of ordinary skill in the art at the time of invention, confronted by the same problems as the inventor and with no knowledge of the claimed invention, would have selected the elements from the prior art and combined them in the claimed manner.

III. PERSON OF ORDINARY SKILL IN THE ART

27. It is my understanding that when interpreting the claims of the '259 patent and evaluating whether a claim is anticipated or would have been obvious, I must do so based on the perspective of a person of ordinary skill in the art at the relevant priority date. I have been instructed to assume for the purposes of my opinions that the relevant priority date of the '259 patent is September 15, 2011.

28. I understand that in determining the level of ordinary skill in the art, several factors are considered. Those factors may include: (i) the type of problems

encountered in the art; (ii) prior art solutions to those problems; (iii) the rapidity with which innovations are made; (iv) the sophistication of the technology; and (v) the educational level of active workers in the field. A person of ordinary skill in the art must have the capability of understanding the scientific and engineering principles applicable to the pertinent art.

29. Based on my review of the specification and claims of the '259 patent, it is my opinion that a person of ordinary skill in the art would have had a minimum of a bachelor's degree in electrical engineering, computer engineering, computer science, or a similar field, and at least three years of industry or academic experience in the design, development, and/or implementation of microphone arrays and/or speech signal processing. Work experience could substitute for formal education and additional formal education could substitute for work experience.

30. My conclusions below that the claims of the '259 patent would have been obvious would remain the same even if the priority date, field of endeavor, or level of ordinary skill were slightly different.

31. I meet the above definition of a person of ordinary skill in the art, and did so as of September 2011. Also, I have worked with persons of ordinary skill in the art through my professional and academic experiences, and I have an understanding of their skill level around September 2011.

IV. TECHNOLOGY BACKGROUND

A. Detecting Speech in Audio Signals Was Known.

32. Detecting the presence or absence of speech in an audio signal is commonly referred to as Voice Activity Detection (“VAD”). For decades, VAD has been an important part of many areas of speech processing, including speech coding, speech enhancement, and automatic speech recognition. In the 1990s, several standards were adopted for implementing VAD in phones. For example, the European Telecommunications Standards Institute (ETSI) adopted VAD techniques for mobile phones. (*E.g.*, EX-1013 (ETSI).)

33. Many prior art references describe using voice activity detectors in communication devices. (*See, e.g.*, EX-1009 (Chan) ¶[0101].) For example, Florencio discloses a VAD 306 that “that classifies each input frame as one of three classes: “Speech, Noise, or Not Sure.” (EX-1005 (Florencio) ¶[0044].) Reuss has a VAD 116 that “determines when the headset user is speaking and when the user is silent (i.e., whether the signals 109 and 111 include voice or only noise).” (EX-1014 (Reuss), 8:30-32; *see also id.*, 8:30-50.) Reuss explains that “VAD functions are well known in the telephony literature.” (*Id.*, 8:39-40.) Visser ’447 explains that using VAD “is desirable to perform accurate detection of segments of an audio signal that carry speech information.” (EX-1015 (Visser ’447) ¶[0075].) Strömmer

includes a VAD that determines if a signal contains speech by detecting “speech like qualities.” (EX-1010 (Strömmer) ¶¶[0053].)

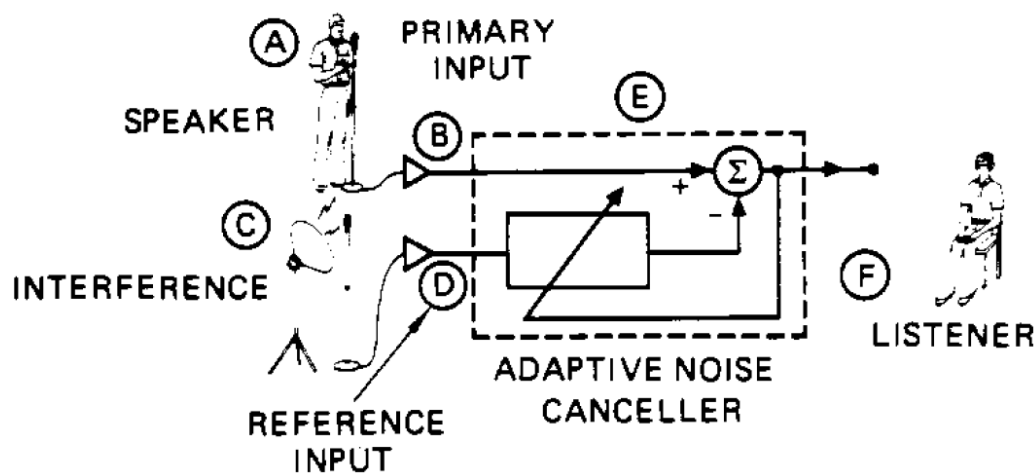
B. Determining an Audio Signal’s Incoming Direction Was Known.

34. Using multiple microphones to determine the direction of an arriving sound has also been known for decades. By the early 2000s, textbooks described various methods for “sound source location,” i.e., determining the direction and location of a sound source. (*E.g.*, EX-1011 (Brandstein), 157-180; EX-1016 (Benesty), 1043-1063.) For example, Strömmer includes a DOA estimation block that estimates the DOA of a sound signal. (EX-1010 (Strömmer) ¶¶[0053], [0060]-[0063].) Florencio includes an SSL block 308 that determines the DOA and location of a sound source from frames classified as speech. (EX-1005 (Florencio) ¶¶[0044], [0048].) Reuss determines the direction of arrival for the strongest noise sources. (EX-1014 (Reuss), 7:24-35.)

C. Reducing Noise via Adaptive Filters Was Known.

35. In signal processing, a filter is a device or process that transforms a signal, e.g., by suppressing unwanted components. (EX-1017 (Stranneby), 17-21.) With digital signals, filtering is accomplished by using computer program code to apply mathematical operations to an input signal, thereby producing filtered output signals. (*Id.*) These mathematical operations typically incorporate parameters (e.g., coefficients or weights) that control the filtering process. (*Id.*)

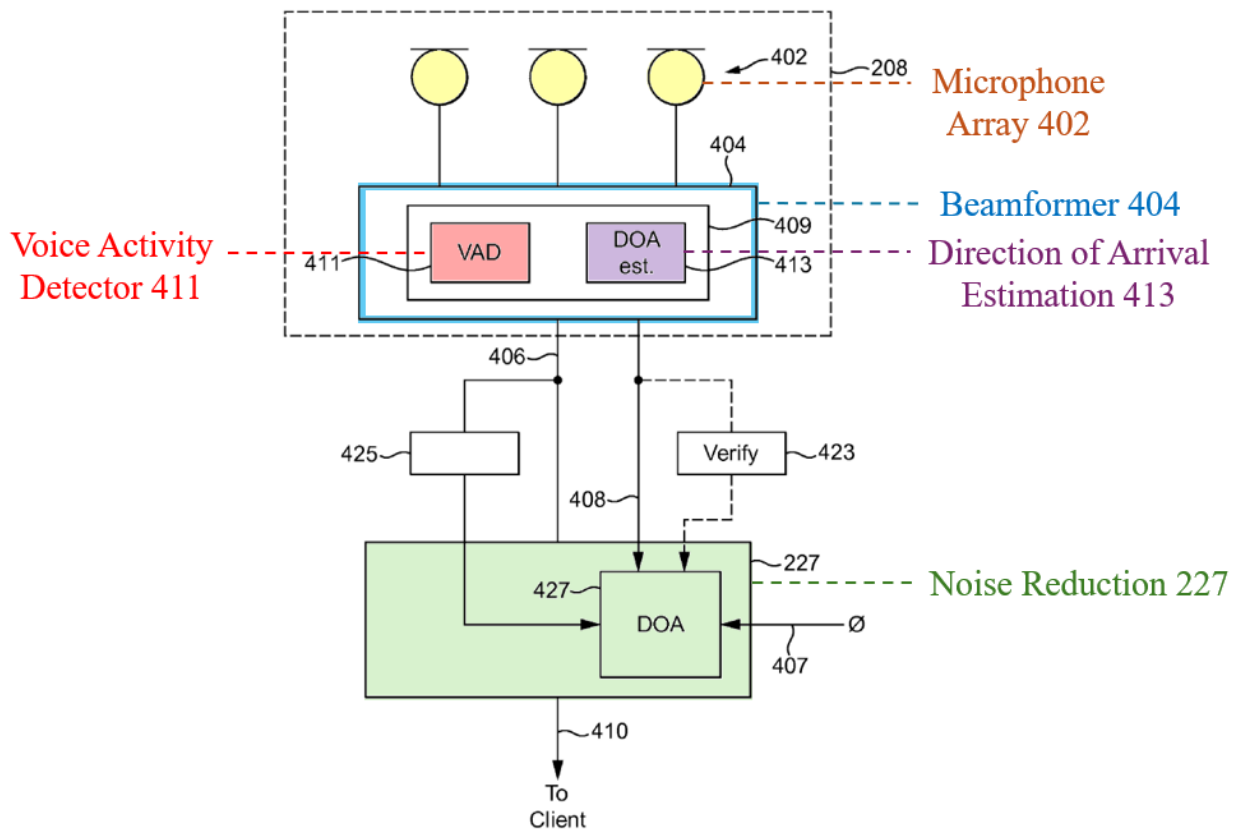
36. An adaptive filter is a type of filter in which the parameters are changed over time to optimize the performance of the filter. (EX-1018 (Comprehensive Dictionary), 14.) The use of adaptive filters to reduce noise in speech signals has been widely known in the art since at least 1975, when Stanford University researchers published a seminal article on the topic including the figure below, which depicts the experiments run by the authors to simulate cockpit noise and reduce that noise with an adaptive noise canceller. (EX-1019 (Widrow), 1704.)



(*Id.*) Since then, many textbooks and other prior art references have described such adaptive filters for noise reduction. (*E.g.*, EX-1011 (Brandstein), 266-276; EX-1014 (Reuss), 10:51-11:26; EX-1004 (Li), Abstract.)

D. Using Voice Activity Detection, DOA Determination, and an Adaptive Filter in the Same Device Was Known.

37. Using these three components together to improve the quality of a sound signal was also known. For example, Strömmer disclosed a system comprising multiple microphones (402), a VAD (411), a DOA estimator (413), and two adaptive filters that reduced noise (beamformer (404) and noise reduction stage (227)):



(EX-1010 (Strömmen), Fig. 4¹.)

¹ Figures in this Declaration may be colored and/or annotated for clarity.

38. As another example, Florencio described a noise reduction system comprising a VAD (306), a sound source localization (“SSL”) unit (308) that determines the direction of arrival (“DOA”), and an adaptive filter (beamformer 310) for reducing noise:

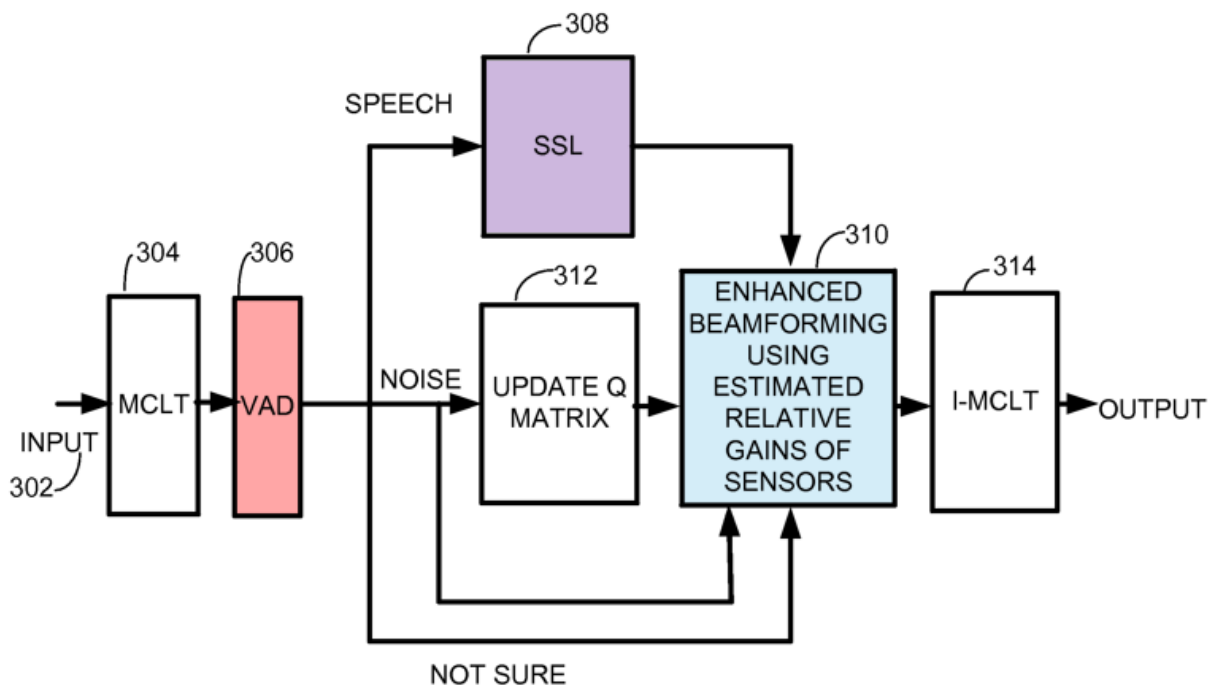


FIG. 3

(EX-1005 (Florencio), Fig. 3.)

39. The beamformers disclosed in Strömmer and Florencio are adaptive filters because they automatically optimize their filtering behavior over time. A beamformer is a spatial filter: it emphasizes signal components coming from a desired direction while suppressing components from other directions. (EX-1005 (Florencio) ¶[0008]; EX-1016 (Benesty), 946.) By focusing a beamformer in the direction

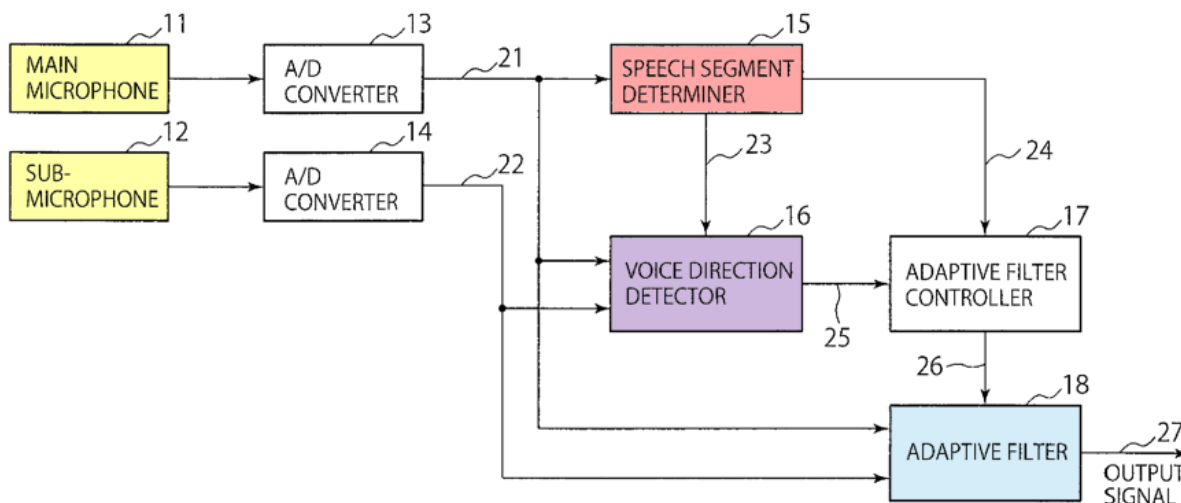
of a person speaking, that person's voice can be emphasized and noise coming from other directions can be filtered out. (*Id.*)

40. Adaptive beamformers, like those disclosed by Strömmer and Florencio, were already well known in the art, as Strömmer acknowledges. (EX-1010 (Strömmer) ¶¶[0008]-[0009].) An adaptive beamformer can change the direction in which it focuses based on a speech signal's estimated direction of arrival. (*Id.*) Thus, the adaptive beamformers disclosed in Strömmer and Florencio are adaptive filters that perform a noise reduction process. (See EX-1011 (Brandstein), 89 (beamformer as “adaptive noise canceler”).)

V. THE '259 PATENT

A. Overview

41. The '259 patent describes a system for reducing noise in audio signals. (EX-1001 ('259 patent), Abstract.) The system includes microphones (11, 12) for capturing sound and three signal processing components: (1) a “speech segment determiner” (i.e., VAD) that determines whether the sound picked up by the microphones contains speech (*id.*, 4:40-43); (2) a “voice direction detector” that determines from which direction a voice sound travels (*id.*, 16:60-65); and (3) an adaptive filter that performs a noise reduction process (*id.*, 22:38-60). The system is shown in Figure 1:



(*Id.*, Fig. 1.)

42. The patent admits that each of the three signal processing components was known.

The speech segment determiner 15 can employ *any speech segment determination techniques*.

(EX-1001 ('259 patent), 4:48-49 (emphasis added).)

There are *several techniques for voice direction detection*. One technique is to detect a voice incoming direction based on a phase difference between the Sound pick-up signals 21 and 22. Another technique is to detect a voice incoming direction based on the difference or ratio between the magnitudes of a sound (the Sound pick-up signal 21) picked up by the main microphone 11 and a sound (the Sound pick-up signal 22) picked up by the sub-microphone 12. The difference and the ratio between the magnitudes of Sounds are referred to as a power difference and a power ratio, respectively. Both factors are referred to as power information, hereinafter.

Whatever the technique is used, the voice direction detector 16 detects a voice incoming direction only when the speech segment determiner 15 determines that a Sound picked up by the main microphone 11 is a speech segment.

(*Id.*, 16:66-17:15 (emphasis added).)

When the relationship $PD1 \geq T$ is established, the adaptive filter controller 17 controls the adaptive filter 18 to perform a regular noise reduction process. The relationship $PD1 \geq T$ indicates that the phase of a voice component carried by a sound pick-up signal 21 obtained based on a sound picked up by the main microphone 11 is more advanced than the phase of a voice component carried by a sound pick-up signal 22 obtained based on a sound picked up by the sub-microphone 12.

In this case, the adaptive filter 18 performs *the regular noise reduction process* to reduce a noise component carried by the sound pick-up signal (a voice signal) 21 using the sound pick-up signal (a noise-dominated signal) 22 to produce the output signal 27. Moreover, in this case, the speech segment determiner 15 detects a speech segment based on the sound pick-up signal 21 obtained based on a sound picked up by the main microphone 11.

(*Id.*, 23:59-24:8 (emphasis added).)

B. Claims

43. I have been asked to consider claims 1-20 in this Declaration. Of those, claims 1, 12, and 13 are independent. Claim 1 is representative and recites “[a] noise reduction apparatus” comprising the following elements:

[a] a speech segment determiner configured to determine whether or not a sound picked up by at least either a first microphone or a second microphone is a speech segment and to output speech segment information when it is determined that the sound picked up by the first or the second microphone is the speech segment;

[b] a voice direction detector configured, when receiving the speech segment information, to detect a voice incoming direction indicating from which

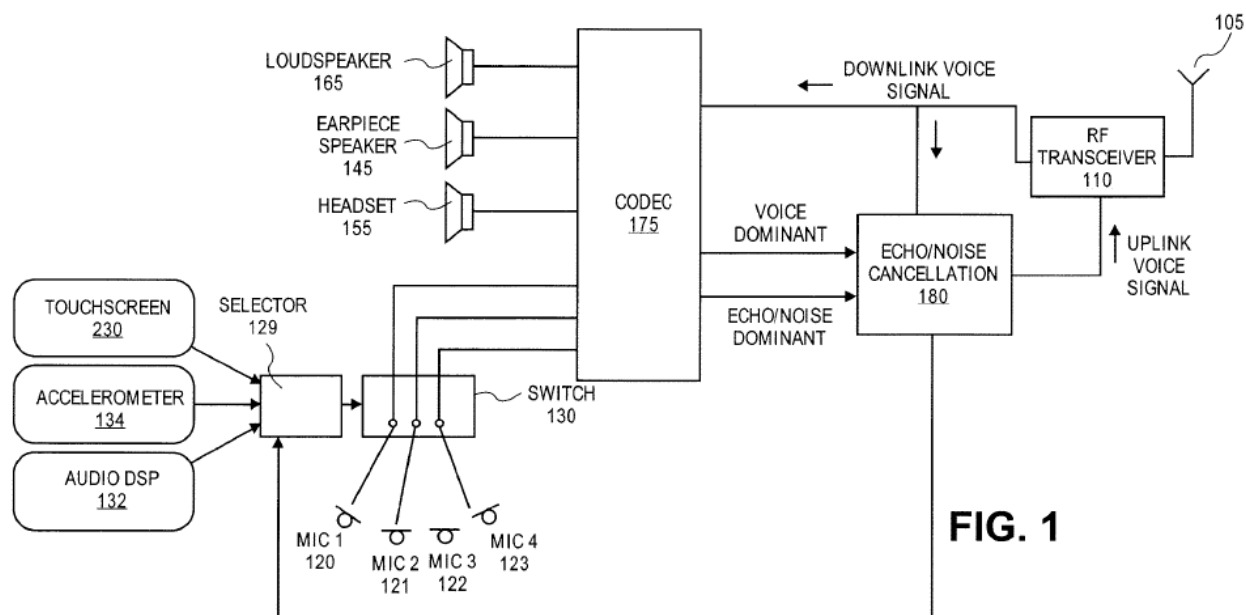
direction a voice sound travels, based on a first sound pick-up signal obtained based on a sound picked up by the first microphone and a second sound pick-up signal obtained based on a sound picked up by the second microphone and to output voice incoming-direction information when the voice incoming direction is detected; and

[c] an adaptive filter configured to perform a noise reduction process using the first and second sound pick-up signals based on the speech segment information and the voice incoming-direction information.

VI. CLAIMS 1-20 OF THE '259 PATENT WOULD HAVE BEEN OBVIOUS.

A. Claims 1-3, 5-6, 8-9, 12-15, and 17-20 Would Have Been Obvious in View of Chen and Li.

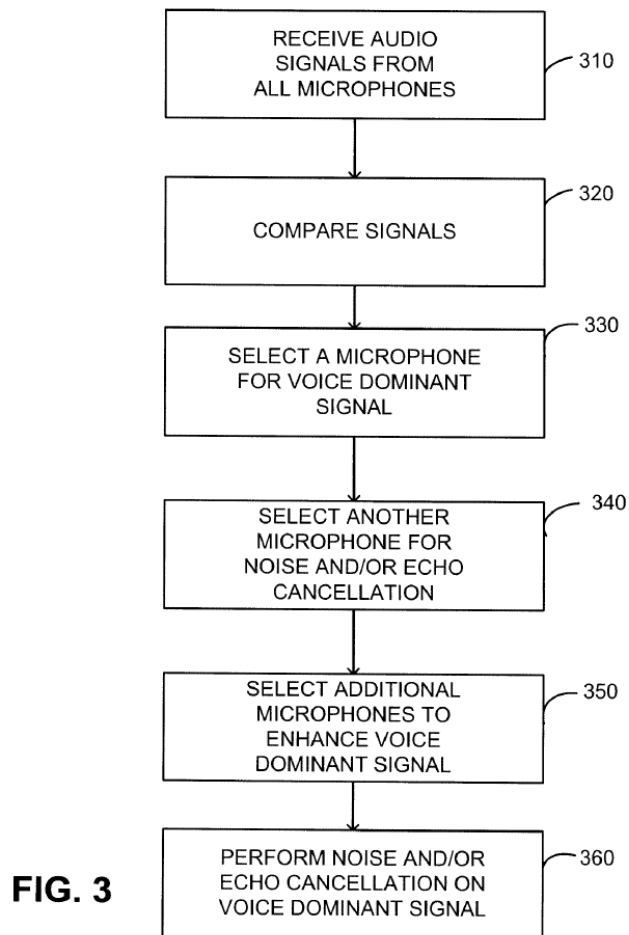
44. Chen discloses a device (e.g., mobile phone) for enhancing voice quality in a signal. (EX-1003 (Chen) ¶[0012].) The device includes several microphones (e.g., 120-123) and a digital signal processor (“DSP”) 132 that “monitors (to compare or analyze) the available microphone signals and provides the results of its analysis to a selector 129.” (*Id.* ¶[0015].)



(*Id.*, Fig. 1.) Selector 129 selects one of the microphones as the “primary” microphone to provide a voice-dominant signal to a codec 175. (*Id.*) The selector also selects a microphone to act as the “secondary” microphone to provide a noise-dominant signal. (*Id.*) The voice-dominant signal is selected based on (a) the signal-to-noise ratio (*id.* ¶[0016]), which is a method of detecting voice activity, and (b) the direction of arrival of the voice signal. (*Id.* ¶[0019] (device can implement “audio tracking or audio beam forming capability using its microphones” to “identify the particular ‘theta’ (angle) at which a target speaker is located, by measuring for the maximum audio signal picked up for the target speaker”).) The voice-dominant and noise-dominant signals are then input to an echo/noise-cancellation controller 180, which “refines the voice dominant signal ... by suppressing noise and cancelling

echo with the assistant of one or more signals from the other selected microphones.”

(*Id.* ¶¶[0022], [0028], Fig. 3.) The process is also shown in Figure 3:



(*Id.*, Fig. 3, ¶[0028].)

45. Li discloses a similar system involving multiple microphone signals—one treated as a “desired voice signal” to be enhanced and a “reference signal”—and an “adaptive filter” used to eliminate noise, enhance the signal-to-noise ratio, and protect speech quality in the desired voice signal. (EX-1004 (Li), Abstract, ¶¶[0010], [0048].) Li’s adaptive filter for noise reduction is shown in Figure 4:

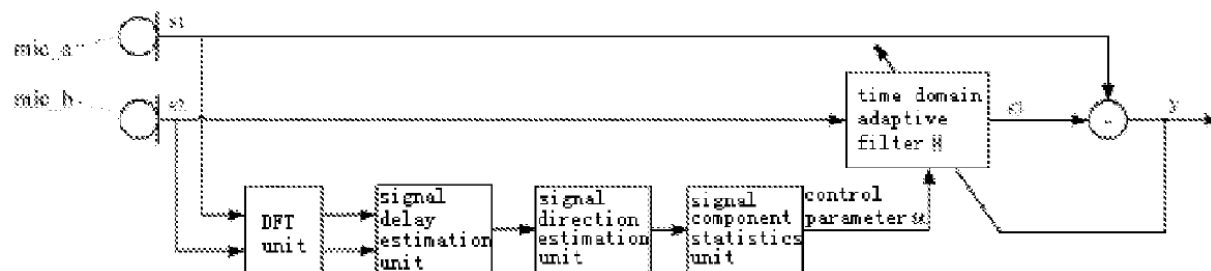


Figure 4

(*Id.*, Fig. 4.)

1. Claim 1

a. Preamble

46. Chen discloses a “noise reduction apparatus” as recited in the preamble because it discloses a handheld device that enhances a “voice dominant signal” by “suppressing ambient noise or canceling echo therein.” (EX-1003 (Chen), Abstract, Figs. 1-4.) An exemplary apparatus is the phone shown in Chen’s Figure 2:

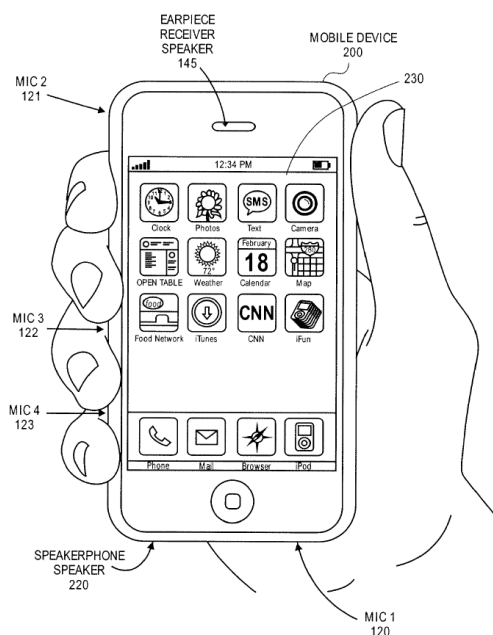


FIG. 2

(*Id.*, Fig. 2.) Thus, Chen discloses the preamble of claim 1.

b. Element 1[a]: Speech Segment Determiner

47. Element 1[a] recites “a speech segment determiner configured to [i] determine whether or not a sound picked up by at least either a first microphone or a second microphone is a speech segment and [ii] to output speech segment information when it is determined that the sound picked up by the first or the second microphone is the speech segment.” Chen discloses or renders obvious this claim element.

48. Chen discloses a digital signal processor 132 that “monitors (to compare or analyze) the available microphone signals and provides the results of its analysis to a selector.” (EX-1003 (Chen) ¶[0015].) The selector selects one of the microphones to act as the “primary” microphone, which provides a “voice dominant signal” and another microphone to serve as the “secondary” microphone, which provides a “noise/echo cancellation dominant signal.” (*Id.*) Chen discloses that the voice-dominant signal may be selected based on, e.g., the “signal-to-noise ratio” (commonly referred to as “SNR”). (*Id.* ¶[0016].) A person of ordinary skill in the art would have understood this disclosure to be a reference to determining whether each microphone signal contains a speech segment. This is because Chen describes selecting a “voice dominant signal” in a mobile phone via SNR, and SNR was a standard way of performing voice activity detection in mobile phones at the time. (*See* EX-1013 (ETSI), 18 (defining SNR), 22 (SNR used in VAD decision).)

49. This claim element also would have been obvious in view of Chen's disclosure. The purpose of Chen's selector 129 is to select the best signal as the voice-dominated signal. A person of ordinary skill would have immediately recognized that using a voice activity detector to identify which of the available microphone signals contained speech, e.g., by using SNR as commonly known, would have aided the selector in choosing the best signal by ensuring that the selected signal in fact contains voice. A VAD also would have helped the selector to select, as the echo/noise-dominant signal, a signal without speech where possible.

50. Chen further discloses that the output of the DSP is provided to the selector. (EX-1003 (Chen) ¶¶[0015] (digital signal processor "monitors (to compare or analyze) the available microphone signals and provides the results of its analysis to a selector 129"), [0034].)

51. Accordingly, Chen discloses or renders obvious a speech segment determiner (e.g., DSP module analyzing SNR) configured to determine whether or not a sound picked up by at least either a first microphone or a second microphone (e.g., microphones 120, 121) is a speech segment and to output speech segment information (e.g., to other DSP modules or selector) when it is determined that the sound picked up by the first or the second microphone is the speech segment.

c. Element 1[b]: Voice Direction Detector

52. Element 1[b] recites “a voice direction detector configured, when receiving the speech segment information, [i] to detect a voice incoming direction indicating from which direction a voice sound travels, based on a first sound pick-up signal obtained based on a sound picked up by the first microphone and a second sound pick-up signal obtained based on a sound picked up by the second microphone and [ii] to output voice incoming-direction information when the voice incoming direction is detected.” Chen discloses or renders obvious this claim element.

53. Chen discloses that the device may:

[I]mplement[] audio tracking or audio beam forming capability using its microphones, to identify the particular “theta” (angle) at which a target speaker is located, by measuring for the maximum audio signal picked up for the target speaker. Thus, as a target speaker is moving around while talking, the tracking/beam forming capability may actively track the strongest signal by switching amongst multiple microphones situated in a microphone array so as to always select the “best” of the available microphone signals as the voice dominant signal.

(EX-1003 (Chen) ¶[0019].) A person of ordinary skill in the art would have understood that this angle identification (i.e., DOA determination) would be performed by Chen’s DSP 129 or processor 704, and Chen discloses providing the result of that analysis to the selector so that it can choose the best available microphone signal.

54. Thus, Chen discloses or renders obvious a voice direction detector configured, when receiving the speech segment information (e.g., speech present based on SNR ratio), to detect a voice incoming direction (e.g., angle) indicating from

which direction a voice sound travels, based on a first sound pick-up signal obtained based on a sound picked up by the first microphone and a second sound pick-up signal obtained based on a sound picked up by the second microphone (e.g., signal obtained based on sound picked up by the microphones 120, 121) and to output voice incoming-direction information when the voice incoming direction is detected (e.g., to the selector).

d. Element 1[c]: Adaptive Filter

55. Element 1[c] recites “an adaptive filter configured to perform a noise reduction process using the first and second sound pick-up signals based on the speech segment information and the voice incoming-direction information.” Chen and Li together disclose or render obvious this claim element.

56. Chen discloses that an “echo/noise cancellation controller 180 refines the voice dominant signal received from the CODEC 175 by suppressing noise and canceling echo with the assistance of one or more signals from the other selected microphones.” (EX-1003 (Chen) ¶[0022]; *id.* ¶¶[0028], [0034].) The “output of the echo/noise cancellation controller 180 provides an enhanced voice signal[.]” (*Id.* ¶[0022]) Thus, Chen discloses using the voice-dominant signal and the echo/noise-dominant signal to perform echo and noise cancellation. (*Id.*, Figs. 1, 3, Abstract.) Chen does not expressly refer to its noise cancellation module as an “adaptive filter.”

57. Li describes an improved adaptive filter for noise reduction in “[w]ireless mobile communication ... devices.” (EX-1004 (Li) ¶[0002].) The device includes two microphones (mic_a and mic_b) that collect acoustic signals s_1 and s_2 . (*Id.* ¶[0048].) Signal s_1 is treated as a “desired voice signal” and s_2 is treated as a “reference signal.” (*Id.*) Those signals are fed into an adaptive filter, which produces a “noise reduced signal y .” (*Id.*) An example of Li’s system is shown in Figure 2:

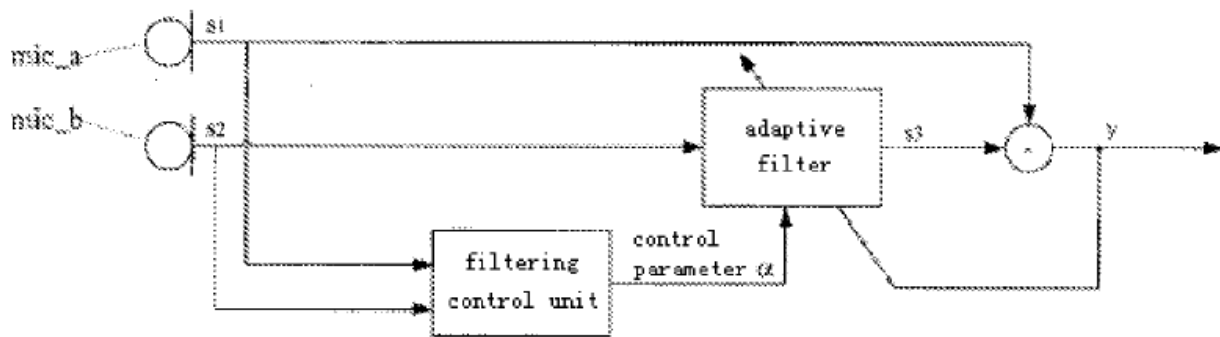


Figure 2

(*Id.*, Fig. 2, ¶¶[0010], [0036], [0045]-[0048]; *id.*, claim 1 (reciting a “noise reduction method ... comprising ... an adaptive filter”).) Li confirms that adaptive filters for noise reduction were widely known in the art. (*Id.* ¶[0045] (“typically, noise reduction is implemented using an adaptive filter”).)

58. When combined with Chen, Li’s adaptive filter would perform a noise reduction process using the first and second sound pick-up signals based on the

speech segment information (e.g., the presence or absence of speech) in two ways. First, as I discuss above, speech segment information is used by Chen to select which microphone signal will be the voice-dominant signal—or “desired voice signal” in Li—to be enhanced by Li’s adaptive filter. Second, Li’s adaptive filter changes based on the presence or absence of speech. For example, where no speech is detected and the microphone signals contain only noise, the adaptive filter performs a certain way. (EX-1004 (Li) ¶¶[0054], [0064].) When both signals contain speech, the adaptive filter “stops updating of weights of the filter ..., thereby protecting speech in the desired speech signal[.]” (*Id.* ¶[0055]; *id.* ¶[0020] (adaptive filter updated fast “when there is only noises” and slow “when there is target signals,” i.e., voice).) Thus, the adaptive filtering would be “based on” the speech segment information.

59. Li’s adaptive filter is also configured to perform a noise reduction process using the first and second sound pick-up signals based on the voice incoming-direction information. First, as I discuss above, the voice incoming direction information is used by Chen to select the microphone signal that will be the voice-dominant signal—or “desired voice signal” in Li—to be enhanced by Li’s adaptive filter. Second, Li’s adaptive filter uses the “incidence angle” of the incoming voice signals to control the adaptive filter. (EX-1004 (Li) ¶[0069], claims 1-6, 10, 11.)

60. Accordingly, Li discloses or renders obvious an adaptive filter as recited in element 1[c].

e. Motivation to Combine

61. A person of ordinary skill would have been motivated to incorporate Li's improved adaptive filter into Chen's echo/noise cancellation stage, for many reasons.

62. First, Chen discloses performing noise reduction to enhance and refine a voice-dominant signal using a noise/echo-dominant signal (*e.g.*, EX-1003 (Chen) ¶[0015], claim 7), but does not provide details of that noise reduction stage. Thus, a person of ordinary skill would have been motivated to look to other references, such as Li, to understand how to implement the noise reduction step. Li describes doing so using an adaptive filter.

63. Second, Li's adaptive filter eliminates noise, enhances SNR, and protects speech quality. (EX-1004 (Li) ¶[0010].) A person of ordinary skill in the art would have been motivated to modify Chen's system to include Li's adaptive filter to obtain the benefits Li describes.

64. Third, as Li recognizes, the use of such adaptive filters for noise reduction was widely known. (EX-1004 (Li) ¶[0045]; *see* ¶¶35-36, above.) Thus, implementing Li's adaptive filter in Chen's system would be the simple substitution of

one known element (Li's adaptive filter for noise reduction) for another known element (Chen's noise reduction stage) to obtain predictable results (improved noise reduction). The combination would also reflect using a known technique (Li's adaptive filter) to improve a similar device and method (Chen's device) in the same way (to enhance signal quality). It would also apply a known technique (Li's adaptive filtering) to a known device and method (Chen's) that is ready for improvement and yield predictable results (improved noise reduction).

65. A person of ordinary skill would have reasonably expected success in modifying Chen in this way. Doing so would have been trivial, involving configuring Chen's device with an adaptive filter as was widely known. Further, Li describes implementing its improved adaptive filter in a very similar system. (EX-1003 (Chen), Abstract; EX-1004 (Li) ¶[0002].)

66. Accordingly, for at least the reasons I discuss above, Chen and Li disclose or render obvious each limitation of claim 1, and therefore render claim 1 obvious as a whole.

2. Claim 13

67. Independent claim 13 recites "[a] noise reduction method" that comprises steps corresponding to the functions of the units recited in claim 1. Chen and Li disclose noise reduction methods. (*See generally* EX-1003 (Chen); EX-1004 (Li).)

68. I have compared the elements in claim 13 to the elements in claim 1. As shown, the steps recited in claim 13 mirror the elements recited in claim 1. The table below shows how the elements correspond. As I discuss above for claim 1, Chen and Li disclose or render obvious the elements recited in claim 13. (See ¶¶46-66, above.)

	Claim 13		Claim 1
13[pre]	A <i>noise reduction</i> method comprising the steps of:	1[pre]	A <i>noise reduction</i> apparatus comprising:
13[a]	<i>determining whether or not a sound picked up by at least either a first microphone or a second microphone is a speech segment;</i>	1[a]	a speech segment determiner configured to <i>determine whether or not a sound picked up by at least either a first microphone or a second microphone is a speech segment</i> and to output speech segment information <u>when it is determined that the sound picked up by the first or the second microphone is the speech segment;</u>
13[b]	<i>detecting a voice incoming direction indicating from which direction a voice sound travels, based on a first sound pick-up signal obtained based on a sound picked up by the first microphone and a second sound pick-up signal obtained based on a sound picked up by the second microphone, when it is <u>determined that the sound picked up by the first or the second microphone is the speech segment;</u> and</i>	1[b]	a voice direction detector configured, when receiving the speech segment information, to <i>detect a voice incoming direction indicating from which direction a voice sound travels, based on a first sound pick-up signal obtained based on a sound picked up by the first microphone and a second sound pick-up signal obtained based on a sound picked up by the second microphone</i> and to output voice incoming-direction

	Claim 13		Claim 1
			information when the voice incoming direction is detected; and
13[c]	<i>performing a noise reduction process using the first and second sound pick-up signals based on speech segment information indicating that the sound picked up by the first or the second microphone is the speech segment and voice incoming-direction information indicating the voice incoming direction.</i>	1[c]	<i>an adaptive filter configured to perform a noise reduction process using the first and second sound pick-up signals based on the speech segment information and the voice incoming-direction information.</i>

69. Accordingly, for at least the reasons I discuss above, Chen and Li render claim 13 obvious as a whole.

3. Claims 2 and 14

70. Claim 2 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on a phase difference between the first and second sound pick-up signals.” Claim 14 depends from claim 13 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals.” Li discloses or renders obvious these claim elements.

71. I note that the ’259 patent explains that the “phase difference” refers to the “amount of delay” between the audio signal being detected by the first and sec-

ond microphones. (EX-1001 ('259 patent), 18:41-47.) Specifically, the phase difference in radians at a given frequency is simply the time delay in seconds multiplied by the frequency in radians per second.

72. Li discloses “calculating phase differences between various frequency bins or sub-bands of the microphone array signals” to calculate relative time delays, and then using those relative time delays in a “signal direction estimation unit for calculating incidence angles of the microphone array signals.” (EX-1004 (Li) ¶[0030]; *id.* ¶[0053] (calculate incidence angles of target speech signals based on “phase differences”).) Thus, Li discloses a voice direction detector that detects the voice incoming direction (incidence angles of target speech signals) based on a phase difference between the first and second sound pick-up signals.

73. A person of ordinary skill in the art would have been motivated to combine the teachings of Chen and Li, and to determine the incoming voice direction at least in part using phase differences (as taught by Li) for several reasons.

74. First, Chen discloses determining the incoming voice direction “by measuring for the maximum audio signal.” (EX-1003 (Chen) ¶[0019].) However, Li teaches that, to enable its improved noise reduction system, the incoming voice direction can be calculated based on phase differences. (EX-1004 (Li) ¶¶[0053], [0030].) Li uses this method to determine the target speech signals within an “angle of protection,” which is a component of Li’s adaptive noise reduction filter. (*Id.*)

Thus, a person of ordinary skill would have understood that implementing Li's adaptive noise reduction filter (as I describe above for claim 1, *see* ¶¶55-66, above) would include Li's voice-angle determination methodology, which is based on phase differences. A person of ordinary skill would have immediately recognized that Li's determination of "incident angles" for speech and noise signals could be used by Chen's processor to help determine the "angle at which a target speaker is located."

75. Second, while Chen discloses determining the angle at which the target speaker is located by measuring for "maximum power," Chen also teaches that the "highest power or loudest signal is not necessarily the appropriate choice for providing the voice dominant signal." (EX-1003 (Chen) ¶[0016].) A person of ordinary skill would have understood that determining the incoming voice direction based on both power (as taught by Chen) and phase difference (as taught by Li) would lead to a more accurate identification of the direction of arrival and thus improved selection of the "best" microphone signal to use as the voice-dominant signal.

76. Third, a person of ordinary skill would have understood that the use of phase-based DOA techniques may be beneficial in some circumstances but magnitude-based DOA techniques may be beneficial in others, and therefore would have been motivated to use both techniques. (EX-1012 (Visser), 28:53-29:34; *See* ¶¶136-145, below.)

77. Fourth, determining the direction of arrival using phase differences (as taught by Li) in Chen's system would merely represent the addition of a known element (determining DOA based on phase differences) to another known element (Chen's DOA determination) to obtain predictable results (more accurate DOA determination). The combination would also reflect using a known technique (determining DOA based on phase differences) to improve a similar device and method (Chen's) in the same way (to identify and enhance desired signals). It would also reflect applying a known technique (DOA based on phase differences) to a known device and method (Chen's) that is ready for improvement and yields predictable results (improved DOA determination for improved signal selection and noise reduction).

78. A person of ordinary skill would have reasonably expected success in using phase-based DOA (as described in Li) in Chen's device because phase-based DOA techniques were widely known, routinely implemented, and Li teaches implementing such techniques on a similar device.

79. Accordingly, for at least the reasons I discuss above, Chen and Li disclose or render obvious each limitation of claims 2 and 14, and therefore render claims 2 and 14 obvious as a whole.

4. Claims 3 and 15

80. Claim 3 depends from claim 2 and further recites that “the adaptive filter performs the noise reduction process to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first sound pick-up signal has a more advanced phase than the second sound pick-up signal whereas the adaptive filter performs the noise reduction process to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up signal when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal.” Claim 15 depends from claim 14 and further recites that “the noise reduction process is performed to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first sound pick-up signal has a more advanced phase than the second sound pick-up signal whereas the noise reduction process is performed to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up signal when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal.” Chen and Li render obvious these claim elements.

81. Chen discloses selecting, from among multiple microphone signals, one to serve as the voice-dominant signal and one to provide the noise/echo-dominant signal. (EX-1003 (Chen) ¶[0015].) Chen and Li both describe using the second (e.g., noise/echo-dominant) signal to reduce noise in the first (e.g., voice-dominant)

signal. Chen also discloses identifying the best microphone and switching between microphones so as to always select the best microphone, e.g., while the target speaker is moving around. (*Id.* ¶¶[0019], [0025].) As I discuss for claim 2 above, Li teaches identifying the direction to the target speaker using phase differences. (See ¶¶70-79, above.)

82. A person of ordinary skill in the art would have understood that the Chen-Li combination using phase-based DOA satisfies this claim limitation. With reference to Chen's Figure 2 (shown below), consider a situation in which the device initially determines via phase-based DOA that the incoming voice direction is from a direction aligned with MIC 1 (e.g., the talker is closest to MIC 1 and therefore MIC 1's signal has a more advanced phase), so Chen's selector 129 selects MIC 1 as the voice-dominant signal and MIC 2's signal as the echo/noise-dominant signal.

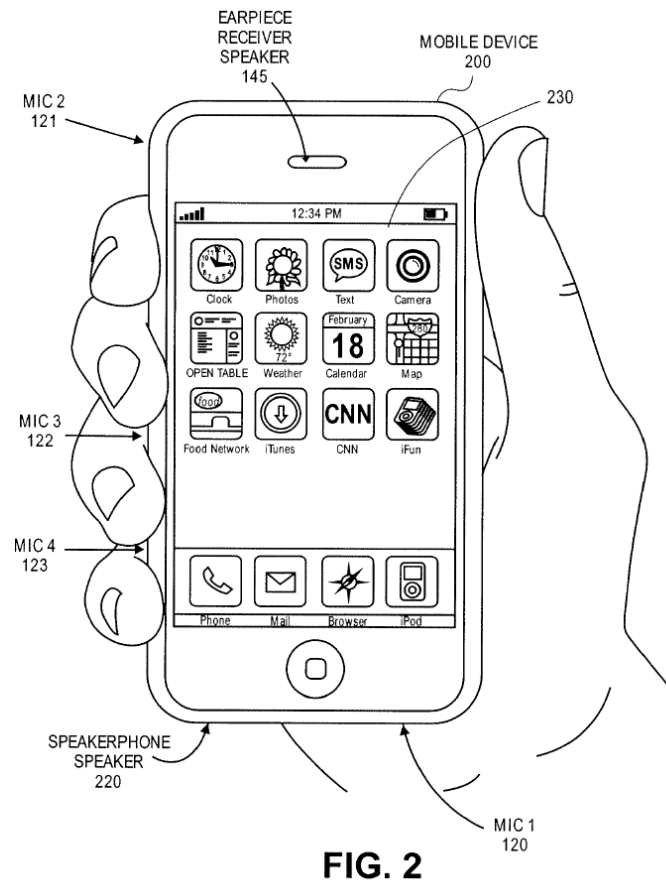


FIG. 2

(EX-1003 (Chen), Fig. 2.)

83. If the target speaker then moves around the device (or changes the orientation of the device) such that the incoming voice direction is from a direction aligned with MIC 2 (e.g., the talker is closest to MIC 2 and therefore MIC 2's signal has the more advanced phase), then Chen's selector would switch so that MIC 2's signal is the voice-dominant signal and MIC 1 is the echo/noise-dominant signal. (EX-1003 (Chen) ¶¶[0019] (Chen switches microphones as the "target speaker is moving around talking" so as to always select the best microphone signal as the

voice-dominant signal), [0004] (switching of primary microphone may “occur dynamically” when the talker “moves from one side of the device to another”).)

84. Thus, Chen and Li render obvious that the adaptive filter performs the noise reduction process to reduce a noise component carried by the first sound pick-up signal (e.g., MIC 1 signal) using the second sound pick-up signal (e.g., MIC 2 signal) when the first sound pick-up signal has a more advanced phase than the second sound pick-up signal (e.g., the talker is closer to MIC 1 than MIC 2) whereas the adaptive filter performs the noise reduction process to reduce a noise component carried by the second sound pick-up signal (e.g., MIC 2 signal) using the first sound pick-up signal (e.g., MIC 1 signal) when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal (e.g., the talker is closer to MIC 2 than MIC 1).

85. Accordingly, for at least the reasons I discuss above, Chen and Li disclose or render obvious each limitation of claims 3 and 15, and therefore render claims 3 and 15 obvious as a whole.

5. Claims 5 and 17

86. Claim 5 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on magnitudes of the first and second sound pick-up signals.” Claim 17 depends from claim 14 and further recites that “the voice incoming direction is detected based on magnitudes of the first

and second sound pick-up signals.” Chen discloses or renders obvious these claim elements.

87. Chen discloses detecting the voice incoming direction (the angle at which a target speaker is located) “by measuring for the maximum audio signal picked up for the target speaker.” (EX-1003 (Chen) ¶[0019].) A person of ordinary skill in the art would have understood Chen’s reference to “maximum audio signal” to refer to the “magnitude” (or “power”) of the signals. Moreover, Chen discloses choosing the best microphone signal based on, among other things, which signal is the “highest powered or loudest signal,” and which has the highest “signal-to-noise ratio.” (*Id.* ¶[0016].)

88. Accordingly, for at least the reasons I discuss above, Chen and Li disclose or render obvious each limitation of claims 5 and 17, and therefore render claims 5 and 17 obvious as a whole.

6. Claims 6 and 18

89. Claim 6 depends from claim 5 and further recites that “the adaptive filter performs the noise reduction process to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first sound pick-up signal has a greater magnitude than the second sound pick-up signal whereas the adaptive filter performs the noise reduction process to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up signal when

the second sound pick-up signal has a greater magnitude than the first sound pick-up signal.” Claim 18 depends from claim 17 and further recites that “the noise reduction process is performed to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first sound pick-up signal has a greater magnitude than the second sound pick-up signal whereas the noise reduction process is performed to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up signal when the second sound pick-up signal has a greater magnitude than the first sound pick-up signal.” Chen discloses or renders obvious these claim elements.

90. As I discuss above for claim 5, Chen discloses or renders obvious determining the voice signals’ direction of arrival based on the signals’ magnitudes. Chen further discloses selecting, as the voice-dominant microphone signal, the microphone signal that has the greater magnitude (e.g., the “highest powered,” “loudest,” or highest “signal-to-noise ratio”). (EX-1003 (Chen) ¶¶[0015]-[0016].) And, the device dynamically switches which microphone signal is the voice-dominant signal and which is the echo/noise-dominant signal. (*Id.* ¶¶[0004], [0019], [0026].)

91. Thus, Chen discloses or renders obvious that the noise reduction process reduces a noise component carried by the first sound pick-up signal (e.g., MIC 1 signal) using the second sound pick-up signal (e.g., MIC 2 signal) when the first sound pick-up signal has a greater magnitude than the second sound pick-up signal

(e.g., when MIC 1 has higher power or loudest signal and is the voice-dominant signal) whereas the noise reduction process reduces a noise component carried by the second sound pick-up signal (e.g., MIC 2 signal) using the first sound pick-up signal (e.g., MIC 1 signal) when the second sound pick-up signal has a greater magnitude than the first sound pick-up signal (e.g., when MIC 2 has higher power or loudest signal is therefore the voice-dominant signal).

92. Accordingly, for at least the reasons I discuss above, Chen and Li disclose or render obvious each limitation of claims 6 and 18, and therefore render claims 6 and 18 obvious as a whole.

7. Claims 8 and 19

93. Claim 8 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” Claim 19 depends from claim 14 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” Thus, claims 8 and 19 recite that voice incoming direction is based on a phase difference (as in claims 2 and 14) and magnitude (as in claims 5 and 17). Chen and Li together disclose or render obvious these claim limitations.

94. As I discuss above, Li discloses determining the voice incoming direction based on a phase difference between the signals. (*See* ¶¶70-79, above.) Chen discloses determining the voice incoming direction based on the magnitudes of the signals. (*See* ¶¶86-88, above.) A person of ordinary skill in the art would have been motivated to incorporate Li's method of determining the voice incoming direction based on phase difference into Chen's device and would have reasonably expected success in doing so for the reasons I discuss above for claim 2. (*See* ¶¶70-79, above.) For example, using both phase difference and magnitudes would have improved Chen's direction determination, would have facilitated the use of Li's adaptive filter, and would have been consistent with Chen's teaching that the selection of the best microphone signal should not be based on any one factor, but on a combination of various factors that will enable the selector to make the best decision. (EX-1003 (Chen) ¶[0016].) Moreover, it would have allowed the device to select or weigh the appropriate DOA technique depending on the circumstances. (EX-1012 (Visser), 28:53-29:34; *see* ¶¶136-145, below.)

95. Accordingly, for at least the reasons I discuss above, Chen and Li disclose or render obvious each limitation of claims 8 and 19, and therefore render claims 8 and 19 obvious as a whole.

8. Claims 9 and 20

96. Claim 9 depends from claim 1 and further recites that “the speech segment determiner detects the speech segment based on the first sound pick-up signal when the first sound pick-up signal has a more advanced phase than the second sound pick-up signal whereas the speech segment determiner detects the speech segment based on the second sound pick-up signal when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal.” Claim 20 depends from claim 19 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” In other words, claims 9 and 20 recite detecting speech based on the sound signal from the microphone nearest to the person talking. Chen discloses or renders obvious these claim elements.

97. In a two-microphone system, Chen describes analyzing both of the microphone signals for speech (e.g., signal-to-noise ratio). (EX-1003 (Chen) ¶¶[0015]-[0016].) Thus, it detects speech based on the first sound pick-up signal (e.g., MIC 1 signal) when the first sound pick-up signal has a more advanced phase than the second sound pick-up signal (e.g., the talker is closer to MIC 1) whereas it detects speech based on the second sound pick-up signal (e.g., MIC 2 signal) when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal (e.g., the talker is closer to MIC 2).

98. Accordingly, for at least the reasons I discuss above, Chen and Li disclose or render obvious each limitation of claims 9 and 20, and therefore render claims 9 and 20 obvious as a whole.

9. Claim 12

99. Independent claim 12 recites “[a]n audio input apparatus” that comprises several units, including those recited in claim 1. Chen and Li each disclose an audio input apparatus. (*See generally* EX-1003 (Chen); EX-1004 (Li).)

100. I have compared the elements in claim 12 to the elements in claim 1. As shown, several of the elements recited in claim 12 mirror the elements recited in claim 1. The table below shows how the elements correspond. As I discuss above for claim 1, Chen and Li disclose or render obvious the corresponding elements recited in claim 12. (*See* ¶¶46-66, above.)

	Claim 12		Claim 1
12[pre]	An audio input apparatus comprising:	1[pre]	A noise reduction apparatus comprising:
12[a]	a first face and an opposite second face that is apart from the first face with a specific distance;		
12[b]	a first microphone and a second microphone provided on the first face and the second face, respectively;		
12[c]	<i>a speech segment determiner configured to determine whether or not a sound picked up by at least either the first mi-</i>	1[a]	<i>a speech segment determiner configured to determine whether or not a sound picked up by at least either a first mi-</i>

	Claim 12		Claim 1
	<i>crophone or the second microphone is a speech segment and to output speech segment information when it is determined that the sound picked up by the first or the second microphone is the speech segment;</i>		<i>crophone or a second microphone is a speech segment and to output speech segment information when it is determined that the sound picked up by the first or the second microphone is the speech segment;</i>
12[d]	<i>a voice direction detector configured, when receiving the speech segment information, to detect a voice incoming direction indicating from which direction a voice sound travels, based on a first sound pick-up signal obtained based on a sound picked up by the first microphone and a second sound pick-up signal obtained based on a sound picked up by the second microphone and to output voice incoming-direction information when the voice incoming direction is detected; and</i>	1[b]	<i>a voice direction detector configured, when receiving the speech segment information, to detect a voice incoming direction indicating from which direction a voice sound travels, based on a first sound pick-up signal obtained based on a sound picked up by the first microphone and a second sound pick-up signal obtained based on a sound picked up by the second microphone and to output voice incoming-direction information when the voice incoming direction is detected; and</i>
12[e]	<i>an adaptive filter configured to perform a noise reduction process using the first and second sound pick-up signals based on the speech segment information and the voice incoming-direction information.</i>	1[c]	<i>an adaptive filter configured to perform a noise reduction process using the first and second sound pick-up signals based on the speech segment information and the voice incoming-direction information.</i>

101. Thus, claim 12 adds two claim elements: element 12[a] recites “a first face and an opposite second face that is apart from the first face with a specific dis-

tance” and element 12[b] recites “a first microphone and a second microphone provided on the first face and the second face, respectively.” Chen discloses these claim elements.

102. In one embodiment, Chen discloses that MIC 1 is on a first face (yellow) and MIC 2 is on a second face (blue):

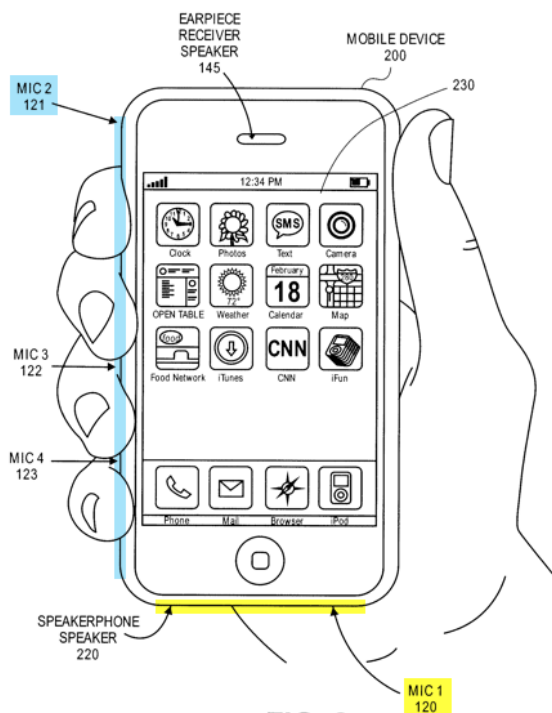


FIG. 2

(EX-1003 (Chen), Fig. 2.) However, Chen teaches that these microphones may be “located anywhere on the device.” (*Id.* ¶[0024]; *id.* ¶[0028] (microphones located “at various locations on the device”).) Thus, Chen discloses or renders obvious microphones on opposing faces (e.g., front and back, top and bottom, left and right sides). Moreover, Chen discloses that, where the device is used to record an interview, one microphone may be selected as the one “most directly facing the interviewee” and another microphone may be facing the interviewer. (*Id.* ¶[0018].) This

would, at a minimum, suggest to a person of ordinary skill in the art placing one microphone on the back of the device and a second microphone on the front. A person of ordinary skill would have been motivated to do so to increase the quality of microphone signals during interviews.

103. Accordingly, for at least the reasons I discuss above, Chen and Li disclose or render obvious each limitation of claim 12, and therefore render claim 12 obvious as a whole.

B. Claims 1-3, 5-6, 8-9, 12-15, and 17-20 Would Have Been Obvious in View of Chen, Li, and Florencio.

104. As I discuss above, Chen and Li disclose or render obvious each limitation of claims 1-3, 5-6, 8-9, 12-15, and 17-20. (*See* ¶¶44-103, above.) For at least those reasons and the reasons I discuss below, those claims would also have been obvious in further view of Florencio. I incorporate my above discussion of Chen and Li here, and only discuss claim elements for which Florencio is relevant in combination with Chen and Li.

105. Florencio describes a noise reduction system for devices such as mobile phones. (EX-1005 (Florencio) ¶[0019].) The system includes a microphone array to receive sound signals and a voice activity detector (306) to determine whether the signals contain speech. (EX-1005 (Florencio) ¶¶[0020]-[0024], [0044]-[0045], [0048], Figs. 1-7.)

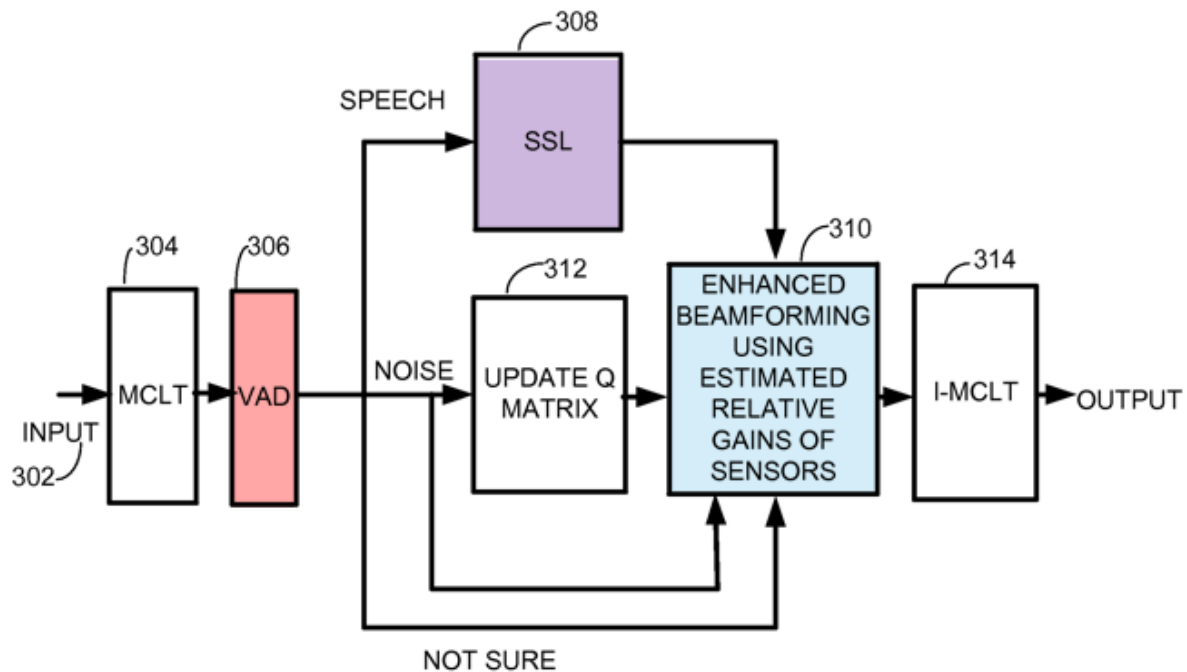


FIG. 3

(*Id.*, Fig. 3.) Sound classified as speech is processed (e.g., by SSL 308) to determine the location of the sound. (*Id.* ¶¶[0048], [0044], Figs. 3, 7.) Sound classified as noise is used to update a Q matrix while sound that is unclassified is sent directly to the beamformer (310). (*Id.*) The beamformer outputs an enhanced signal. (*Id.* ¶¶[0044], [0048], Figs. 3, 7.) A flow diagram is shown below:

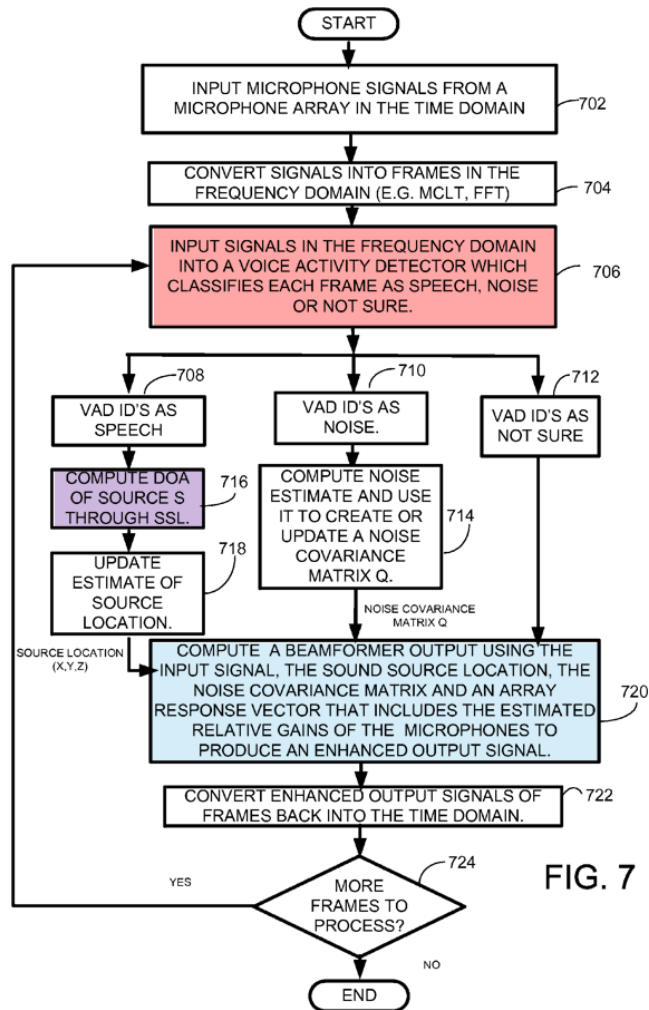


FIG. 7

(*Id.*, Fig. 7.)

1. Claim 1

106. As I discuss elsewhere, Florencio discloses a speech segment determiner (e.g., VAD 306) configured to determine whether or not a sound picked up by the first or second microphone is a speech segment and to output speech segment information (e.g., presence of speech in a signal) when speech is present. (*See* ¶105, above; ¶¶254-256, below.) Florencio also discloses a voice direction detector (e.g.,

SSL 308) configured, when receiving that speech information, to detect a voice incoming direction indicating from which direction a voice sound travels, and to output that information (e.g., to a beamformer). (*Id.*; see ¶¶254-259, below.) And, as I discuss above, Li’s adaptive filter performs noise reduction based on the speech segment information and the voice incoming-direction information.

107. I understand the patent owner may argue that the Chen-Li combination does not disclose or render obvious using VAD and SSL to help select the voice-dominant signal in Chen or enable Li’s adaptive filter. While I disagree, I note that these limitations would also have been obvious in view of Florencio. A person of ordinary skill in the art would have been motivated to modify the Chen-Li combination to include a VAD and SSL unit, as taught by Florencio, for many reasons.

108. First, Chen teaches that the microphone signals are analyzed and compared by a DSP or signal processing module. (EX-1003 (Chen) ¶¶[0015], [0034].) Then, the selector (129 or 732) identifies the “best” of the available microphone signals as the voice-dominant signal. (*Id.* ¶¶[0004], [0019], [0028].) Chen explains that the “highest powered” or “loudest” signal is not necessarily the appropriate choice and that “[v]arious factors ... may be taken into consideration[.]” (*Id.* ¶[0016].) A person of ordinary skill would have recognized that analyzing the signals to determine whether the microphone signals contain speech, as taught by Florencio, would have helped the selector determine which signal should be the voice-

dominant signal. For example, Florencio’s VAD classifies each frame of each signal as either speech, noise, or “not sure.” (EX-1005 (Florencio) ¶[0048].) A person of ordinary skill would have immediately recognized that such a VAD would help Chen’s selector by ensuring that the signal selected as the voice-dominant signal actually contains speech.

109. Second, Chen discloses using “audio tracking or audio beam forming” to identify the angle at which the target speaker is located. (EX-1003 (Chen) ¶[0019].) Consequently, a person of ordinary skill would have been motivated to look to other references, like Florencio, that describe beamforming methods involving DOA determination and its implementation. And, especially in the scenario Chen describes regarding a moving target, both VAD and SSL would help the selector identify the best microphone signal to use at any given time.

110. Third, the use of VADs and SSL algorithms in beamformers was widely known, as the patent admits and many prior art references demonstrate. (See ¶¶32-34, above.) Thus, implementing VAD and SSL (as taught by Florencio) in Chen’s microphone signal analyzer would merely represent the simple addition of known elements (Florencio’s VAD and sound source location functions) to another known element (Chen’s device) to obtain predictable results (a VAD that facilitates determining DOA, which helps identify the direction of the desired signal). The combination would also reflect using a known technique (Florencio’s VAD and SSL) to

improve a similar device (Chen's) in the same way (determining signals with speech and their direction). It would also apply a known technique (VAD and SSL) to a known device (Chen's) that is ready for improvement and yields predictable results (improved ability to select the best signal as the voice-dominant signal by determining whether speech is present and its direction).

111. Fourth, Li's adaptive filter relies on the presence or absence of speech and the "angle of incidence" of the target sound signals to operate effectively. (*E.g.*, EX-1004 (Li), claim 1, ¶¶[0065], [0069], [0020].) Thus, a person of ordinary skill would have immediately recognized that implementing a VAD and SSL unit, as taught by Florencio, in the Chen-Li system would improve the operation of Li's adaptive filter, thus producing a better voice signal.

112. A person of ordinary skill would have reasonably expected success in modifying Chen and Li in this way. Doing so would have been trivial, involving configuring Chen's DSP with revised algorithms that include VAD and SSL units as taught by Florencio.

113. Accordingly, for at least the reasons I discuss above, Chen, Li, and Florencio disclose or render obvious each limitation of claim 1, and therefore render claim 1 obvious as a whole.

2. Claim 13

114. As I discuss above, claim 13 recites a method claim for performing the same limitations as the apparatus of claim 1. (See ¶68, above.) Accordingly, Chen, Li, and Florencio render obvious claim 13 for the reasons I discuss above for claim 1. (See ¶¶46-66, 106-113, above.)

3. Claims 2 and 14

115. Claim 2 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on a phase difference between the first and second sound pick-up signals.” Claim 14 depends from claim 13 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals.” Florencio discloses or renders obvious these claim elements.

116. Florencio discloses using an SSL algorithm “based on time delay of arrival of the signal and maximum likelihood estimation.” (EX-1005 (Florencio) ¶[0044].) Florencio therefore also discloses or renders obvious using the phase differences to determine the DOA as recited in claims 2 and 14. (See ¶¶268-270, below.)

117. Accordingly, for at least the reasons I discuss above, Chen, Li, and Florencio disclose or render obvious each limitation of claims 2 and 14, and therefore render claims 2 and 14 obvious as a whole.

4. Claims 5 and 17

118. Claim 5 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on magnitudes of the first and second sound pick-up signals.” Claim 17 depends from claim 14 and further recites that “the voice incoming direction is detected based on magnitudes of the first and second sound pick-up signals.” Chen and Florencio together render obvious these claim elements.

119. In view of Chen’s disclosure (which I discuss above, *see* ¶¶86-88, above), a person of ordinary skill in the art implementing Florencio’s SSL unit would have been motivated to determine the DOA based at least in part on the magnitude of the signals. Moreover, Florencio discloses that the sound source location and received speech frame are input into a beamforming module which “finds the best output signal to noise ratio[.]” (EX-1005 (Florencio) ¶[0044].) A person of ordinary skill would have understood Florencio’s disclosure to refer the “magnitude” of the signals in each direction. Thus, Chen and Florencio together render obvious that the SSL unit would detect the voice incoming direction “based on magnitudes of the first and second sound pick-up signals.”

120. Accordingly, for at least the reasons I discuss above, Chen, Li, and Florencio disclose or render obvious each limitation of claims 5 and 17, and therefore render claims 2 and 14 obvious as a whole.

5. Claims 3, 6, 8-9, 12, 15, and 18-20

121. As I discuss above, Chen and Li disclose or render obvious the additional claim elements of claims 3, 6, 8-9, 15, and 18-20, and the elements of claim 12 that differ from claim 1. (*See* ¶¶80-103, above.) Accordingly, for at least the reasons I discuss above, Chen, Li, and Florencio disclose or render obvious each limitation of claims 3, 6, 8-9, 12, 15, and 18-20, and therefore render those claims obvious as a whole.

C. Claims 4, 7-8, 12, 16, and 19-20 Would Have Been Obvious in View of (i) Chen, Li, and Visser or (ii) Chen, Li, Florencio, and Visser.

122. Claims 4, 7-8, 16, and 19-20 depend, directly or indirectly, from claims 1 and 13. Furthermore, claim 12 includes limitations that mirror those of claim 1. As I discuss above, Chen, Li, and/or Florencio disclose or render obvious each limitation of claims 1 and 13. (*See* ¶¶46-69, 106-114, above.) In addition, Chen, Li, and/or Florencio disclose or render obvious the additional limitations of claims 8, 19, and 20. (*See* ¶¶93-98, above.) For at least those reasons and the reasons I discuss below, claims 4, 7-8, 12, 16, and 19-20 would also have been obvious in further view of Visser.

123. I incorporate my above discussion of Chen, Li, and Florencio here, and only discuss claim elements for which Visser is relevant in combination with Chen, Li, and/or Florencio.

1. Claims 4 and 16

124. Claim 4 depends from claim 2 and further recites that “when the phase difference is within a predetermined range, the adaptive filter outputs either the first or the second sound pick-up signal without performing the noise reduction process.” Claim 16 depends from claim 14 and further recites the same limitation. Visser discloses or renders obvious these claim elements.

125. Visser discloses a system to reduce noise in audio signals detected by a microphone array. (EX-1012 (Visser), 6:22-31.) Visser explains how its system works when an audio source is “broadside” to its microphone array (i.e., far from an axis of the microphone array):

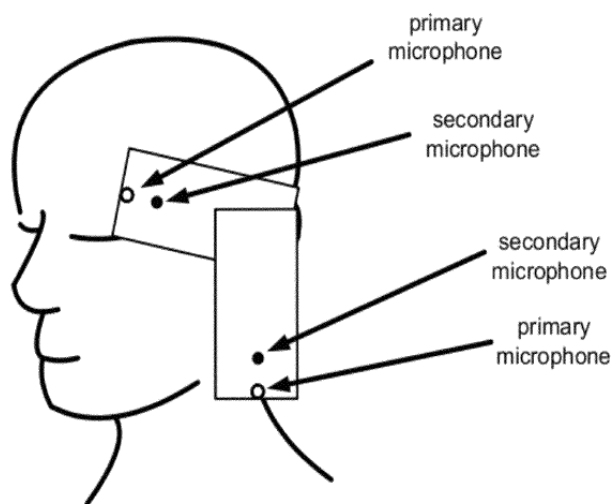
When the handset is held away from the mouth (e.g., in either of the angular holding positions shown in FIG. 2), the ***axis of the microphone array is broadside to the mouth***, and ***effective dual-microphone noise reduction may not be possible***. Use of dual-microphone noise reduction during time intervals in which the handset is held in such a position may result in attenuation of the desired voice signal. For handset mode, a dual-microphone-based scheme typically ***cannot offer consistent noise reduction across a wide range of phone holding positions without attenuating desired speech*** level in at least some of those positions.

(*Id.*, 7:11-21 (emphasis added).)

In a broadside arrangement, the microphone array is oriented relative to the signal source (e.g., a user’s mouth) such that the direction from the center of the array to the source is roughly ***orthogonal*** to the axis of the array. ***Such an arrangement produces two mixtures of desired speech-noise signals that are basically very similar***.

(*Id.*, 7:49-54 (emphasis added).) Figure 2 depicts examples of when a speaker is

broadside to microphones on a mobile phone:



(*Id.*, Fig. 2.)

126. In this scenario, Visser does not perform “dual-microphone noise reduction,” but rather “single-microphone noise reduction.” (EX-1012 (Visser), 7:22-25.) Visser explains that such microphone systems can use “inter-microphone phase difference to determine whether a ... signal originated from within a range of allowable inter-microphone angles or from outside it.” (*Id.*, 5:15-20.) A person of ordinary skill in the art would have understood that Visser discloses using a range of phase differences to determine when the speaker is broadside to the microphone array so that it uses one microphone signal instead of dual-microphone noise reduction. Thus, Visser discloses or renders obvious the additional limitations of claims 4 and 16.

127. A person of ordinary skill in the art would have been motivated to modify the Chen-Li combination to include Visser's single-signal processing option for many reasons.

128. First, Visser explains that when the audio signal is broadside to the array, a dual-microphone noise reduction will attenuate the desired speech signal. (EX-1012 (Visser), 7:14-17.) Chen and Li disclose systems that use dual microphones. Thus, a person of ordinary skill would have been motivated to address such a scenario in the Chen-Li dual microphone system by using one solution Visser describes, e.g., to not use dual-microphone noise reduction and to instead output one of the signals. A person of ordinary skill would have immediately recognized that Visser's solution (e.g., selecting one microphone signal and not performing dual-microphone noise reduction) would improve the Chen-Li combination.

129. Second, the use of SSL techniques and dual- and single-microphone noise reduction were widely known, as Visser and other references demonstrate. (EX-1012 (Visser), 1:51-55; *see* ¶¶34-40, above.) Thus, implementing single-signal noise reduction (as taught by Visser) in the combined Chen-Li SSL and noise reduction system when DOA is broadside would merely represent the simple addition of known elements (Visser's single-microphone selection when broadside) to another known element (Chen's device) to obtain predictable results (skipping dual-microphone noise reduction that may deteriorate the sound signal when the DOA is within

a certain broadside range). The combination would also reflect using a known technique to improve a similar device and method (Chen's) in the same way (improved processing of broadside signals). It would also apply a known technique to a known device (Chen's) that is ready for improvement and yields predictable results (improved signal output for broadside signals).

130. A person of ordinary skill would have reasonably expected success in modifying the Chen-Li combination in this way. Doing so would have been trivial, involving configuring Chen's signal processing with revised algorithms that include Visser's SSL and noise reduction selection method for broadside signals.

131. Accordingly, for at least the reasons I discuss above, (i) Chen, Li, and Visser and/or (ii) Chen, Li, Florencio, and Visser disclose or render obvious each limitation of claims 4 and 16, and therefore render claims 4 and 16 obvious as a whole.

2. Claim 7

132. Claim 7 depends from claim 5 and further recites that "when a power difference that is a difference between magnitudes of the first and second sound pick-up signals is within a predetermined range, the adaptive filter outputs either the first or the second sound pick-up signal without performing the noise reduction process." I note that claim 7 is similar to claim 4 but it uses a "difference of magnitude"

rather than a “difference of phase.” Visser discloses or renders obvious this claim element.

133. As I discuss above for claim 5, Chen discloses using magnitude to determine DOA. (See ¶¶86-88, above.) So does Visser. (EX-1012 (Visser), 6:57-61 (multiple methods for determining DOA, including “phase and/or level (e.g., amplitude, gain, energy) between the channels”).) Visser also discloses that, when a signal is broadside, a single sound pick-up signal should be output rather than using the two-microphone noise reduction process. (See ¶¶124-131, above.)

134. A person of ordinary skill in the art would have understood from Visser that, regardless of whether a phase-based or magnitude-based DOA technique is used, the dual-microphone noise reduction technique should be avoided when the source is broadside. (See ¶¶124-131, above.) Thus, a person of ordinary skill would have known to set a predetermined range of magnitude differences that would indicate that the speaker is broadside. In other words, if the difference in magnitudes of the signals for the microphones was within a certain range, the direction of the speaker is broadside to the microphones. In that scenario, as I discuss above, Visser discloses outputting either microphone signal without performing the dual-microphone noise reduction. (See ¶¶124-131, above.)

135. Accordingly, for at least the reasons I discuss above, (i) Chen, Li, and Visser and/or (ii) Chen, Li, Florencio, and Visser disclose or render obvious each limitation of claim 7, and therefore render claim 7 obvious as a whole.

3. Claims 8 and 19

136. Claim 8 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” Claim 19 depends from claim 14 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” Thus, claims 8 and 19 recite that voice incoming direction is based on a phase difference (as in claims 2 and 14) and magnitude (as in claims 5 and 17). Visser discloses these claim limitations.

137. Visser discloses using both phase differences and magnitude differences for DOA determination. Specifically, Visser discloses:

Examples of SSP operations include *operations that indicate directions of arrival (DOAs) of one or more frequency components of the received multichannel signal, based on differences in phase and/or level (e.g., amplitude, gain, energy) between the channels.*

(EX-1012 (Visser), 6:57-61 (emphasis added).)

As described above, method M100 may be configured to select the channels corresponding to a particular endfire microphone pair according to *DOA information that is based on phase differences* between

channels at different frequencies. *Alternatively or additionally*, method M100 may be configured to *select the channels corresponding to a particular endfire microphone pair according to DOA information that is based on gain differences between channels*. Examples of gain-difference-based techniques for directional processing of a multichannel signal include (without limitation) beam forming, blind source separation (BSS), and steered response power-phase transform (SRP-PHAT). Examples of beam forming approaches include generalized sidelobe cancellation (GSC), minimum variance distortionless response (MVDR), and linearly constrained minimum variance (LCMV) beam formers. Examples of BSS approaches include independent component analysis (ICA) and independent vector analysis (IVA).

(*Id.*, 28:53-29:3 (emphasis added).)

Method M110 may be implemented to select a subset *using phase-difference-based processing as described above at some times, and using gain-difference based processing at other times*, depending on an estimated range of the source (i.e., an estimated distance between source and microphone).

(*Id.*, 29:8-13 (emphasis added).)

Such an implementation of method M110 may be configured to select a subset of channels by combining directional indications from *phase-difference-based and gain-difference-based processing techniques*. For example, such an implementation may be configured to weight the directional indication of a phase-difference-based technique more heavily when the estimated range is small and to weight the directional indication of a gain-difference-based technique more heavily when the estimated range is large. Alternatively, such an implementation may be configured to *select the subset of channels based on the directional indication of a phase difference-based technique when the estimated range is small and to select the subset of channels based on the directional indication of a gain-difference-based technique instead when the estimated range is large*.

(*Id.*, 29:20-34 (emphasis added).) Thus, Visser discloses this limitation.

138. A person of ordinary skill in the art would have been motivated to use both phase-based-differences and magnitude-based-differences to determine DOA in the Chen-Li combination (and the Chen-Li-Florencio combination) for many reasons.

139. First, Visser explains that phase-difference-based DOA techniques “produce good results when the sound source ... [is] close to the microphones (e.g., within one meter), but their performance may fall off at greater source-microphone distances.” (EX-1012 (Visser), 29:4-13.) Thus, a person of ordinary skill would have understood that it would be beneficial to switch between phase-based and magnitude-based DOA techniques “depending on the ... estimated distance between source and microphone.” (*Id.*)

140. Second, Visser describes selecting the microphone signals by combining DOA indications from both phase-based and gain-based processing techniques, for example, by weighting the result of the phase-based technique more heavily when the estimated range is small and the result of the gain-based technique more heavily when the estimated range is large. (EX-1012 (Visser), 29:20-34.) A person of ordinary skill would have been motivated to use both phase-based and magnitude-based DOA in the Chen-Li combination to obtain the performance benefits Visser describes.

141. Third, Chen and Visser both describe methods for determining DOA and reducing noise in mobile phones. (EX-1003 (Chen) ¶¶[0012]; EX-1012 (Visser), Fig. 1, 6:41-49.)

142. Fourth, in view of the discussion in Chen (and Li and Florencio) regarding determining the direction of arrival, a person of ordinary skill would have been motivated to look at other references for improved DOA techniques. Visser describes such a technique and explains its benefits in certain environments.

143. Fifth, determining the voice incoming direction based on both phase differences and magnitude differences would represent the use of a known technique (e.g., Visser's use of both techniques) to improve similar devices (e.g., the Chen-Li DOA unit) in the same way (e.g., improved DOA determination). It also applies a known technique (e.g., Visser's combined DOA approach) to a known device (e.g., Chen's) that is ready for improvement to yield predictable results (e.g., improved DOA determination).

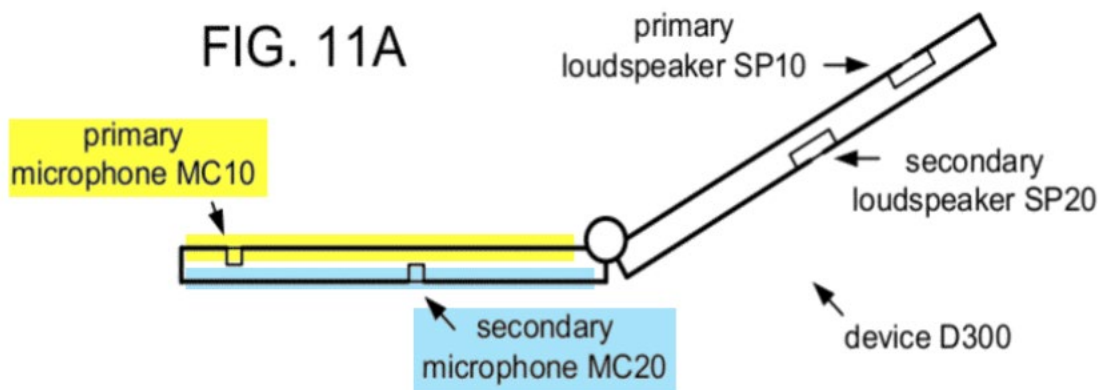
144. A person of ordinary skill would have had a reasonable expectation of success in using both SSL techniques (as taught by Visser) because the techniques individually were generic and well known, and because Visser teaches how to implement the combined approach in a device such as the one Chen describes (e.g., a phone). Thus, implementing the combined DOA approach as Visser teaches would have been within the skill of a person of ordinary skill.

145. Accordingly, for at least the reasons I discuss above, (i) Chen, Li, and Visser and/or (ii) Chen, Li, Florencio, and Visser disclose or render obvious each limitation of claims 8 and 19, and therefore render claims 8 and 19 obvious as a whole.

4. Claim 12

146. As I discuss above (*see* ¶¶100-101, above), claim 12 recites two claim elements in addition to those recited in claim 1: element 12[a] recites “a first face and an opposite second face that is apart from the first face with a specific distance” and element 12[b] recites “a first microphone and a second microphone provided on the first face and the second face, respectively.” Visser discloses these claim elements.

147. Visser discloses a mobile phone having a primary microphone MC10 and a secondary microphone MC20 on opposite faces:



(EX-1012 (Visser), Fig. 11A, 10:31-35.)

148. A person of ordinary skill in the art would have been motivated to implement Visser's microphone orientation in Chen's phone. First, Chen discloses a mobile phone and teaches that the microphones can be placed anywhere on it. (*See* ¶¶99-103, above.) A person of ordinary skill would have been motivated to look at references, such as Visser, to determine where to arrange the microphones in a mobile device to reduce noise.

149. Second, Chen teaches orienting one microphone directly towards an interviewee and another microphone towards the interviewer. (EX-1003 (Chen) ¶[0018].) Visser's microphone placement would provide this desired orientation.

150. Third, in a multi-microphone system to reduce noise, a person of ordinary skill would have been motivated to arrange one microphone facing the anticipated direction of the desired signal (e.g. speaker's mouth) and a second facing a direction of anticipated noise (e.g., facing outward).

151. A person of ordinary skill would have reasonably expected success in implementing Visser's microphone arrangement because microphone placement is a simple design choice, as evidenced by Chen's disclosure that the microphones can be located anywhere on the device.

152. Accordingly, for at least the reasons I discuss above, (i) Chen, Li, and Visser and/or (ii) Chen, Li, Florencio, and Visser disclose or render obvious each limitation of claim 12, and therefore render claim 12 obvious as a whole.

5. Claim 20

153. Claim 20 depends from claim 19 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” As I discuss above (*see* ¶¶96-98, above), Chen discloses or renders obvious the additional claim element of claim 20.

154. Accordingly, for at least the reasons I discuss above, (i) Chen, Li, and Visser and/or (ii) Chen, Li, Florencio, and Visser disclose or render obvious each limitation of claim 20, and therefore render claim 20 obvious as a whole.

D. Claims 10 and 11 Would Have Been Obvious in View of (i) Chen, Li, Kim, and Kleffner or (ii) Chen, Li, Florencio, Kim, and Kleffner.

155. Claims 10 and 11 depend from claim 1. As I discuss above, Chen, Li, and/or Florencio disclose or render obvious each limitation of claim 1. (*See* ¶¶46-66, 106-113, above.) For at least those reasons and the reasons I discuss below, claims 10 and 11 would have been obvious in further view of Kim and Kleffner. I incorporate my above discussion of Chen, Li, and Florencio here.

1. Claim 10

156. Claim 10 depends from claim 1 and further recites that “signals are supplied to the voice direction detector as the first and second sound pick-up signals at a sampling frequency of 24 KHz or higher and signals are supplied to the adaptive

filter as the first and second sound pick-up signals at a sampling frequency of 12 KHz or lower.” Kim and Kleffner disclose or render obvious this claim element.

157. Li discloses using a sampling rate of 8 kHz for its adaptive filter. (EX-1004 (Li) ¶[0052].) Thus, Li teaches the second part of claim 10, i.e., that signals are supplied to the adaptive filter at a sampling frequency of 12 kHz or lower. Neither Chen nor Florencio discuss sampling frequencies for their VAD or DOA determination.

158. Kim discloses processing an audio signal using various sampling frequencies, including 48 kHz. (EX-1006 (Kim), 25, 14.) Kim estimates the direction of arrival information of an audio signal before downsampling the audio signal “in order to maximize the angle resolution of the DOA.” (*Id.*) Downsampling reduces the number of samples in an audio signal. For example, if an original audio signal was sampled at 48kHz (i.e., 48,000 samples per second), downsampling by a factor of 4 reduces the number of samples to 12kHz (i.e., 12,000 samples per second).

159. Kleffner discloses audio signals with multiple sampling frequencies. (EX-1007 (Kleffner), 26.) For example, Kleffner discloses audio signals sampled at 32 or 48 kHz downsampled to 8 kHz. (*Id.*)

160. A person of ordinary skill in the art would have been motivated to use an audio signal with a higher sampling frequency (e.g., 48 kHz) in Chen’s DSP unit to determine the DOA for several reasons.

161. First, although Li teaches using an 8 kHz sampling frequency for the noise reduction stage, Chen does not disclose the sampling frequency used for determining DOA. Thus, a person of ordinary skill would have been motivated to look to references such as Kim and Kleffner to see what sampling frequency to use.

162. Second, Kim discloses processing audio signals with a higher sampling frequency to determine the direction of arrival and then downsampling the audio signal for further processing. (EX-1006 (Kim), 25.) Kim explains that determining the direction of arrival with a higher sampling frequency maximizes the angle of arrival. (*Id.*; *see also id.*, 19 (“a higher sampling rate is better” for DOA).)

163. Third, standard sampling frequencies for music and other audio signals include 48 kHz and 44.1 kHz. Thus, a person of ordinary skill would have been motivated to use a higher sampling frequency, such as 48 kHz. Moreover, a person of ordinary skill would further have been motivated to use 48 kHz because Kim discloses using 48 kHz as a sampling frequency. Kleffner also discloses using 48kHz sampling frequency. (EX-1007 (Kleffner), 26.)

164. Fourth, using a higher frequency such as 48 kHz would merely use a known technique (e.g., Kim’s higher sampling frequency) to improve similar devices (e.g., Chen’s DOA unit) in the same way (e.g., improved accuracy of DOA angle). Moreover, it would merely be the application of a known technique (e.g., Kim’s higher sampling frequency) to a known device (e.g., Chen’s DOA unit) that

is ready for improvement to yield predictable results (e.g., more accurate DOA determination based on higher sampling frequency).

165. Fifth, while using a high sampling frequency improves the DOA determination, a person of ordinary skill would have understood that it is computationally intensive to use a high sampling frequency at every stage in the audio processing. Thus, a person of ordinary skill would have been motivated to downsample the audio signal for the remainder of the processing to reduce the processing requirements in the Chen-Li (and optionally Florencio) combination. Kim and Kleffner both disclose downsampling the audio signal after using the higher sampling frequency. (EX-1006 (Kim), 25; EX-1007 (Kleffner), 26.) And, Li teaches that 8 kHz is sufficient for the noise reduction process. Thus, a person of ordinary skill would have been motivated to downsample the audio signal to a lower sampling frequency (e.g., 8 kHz) for the noise reduction stage.

166. A person of ordinary skill would have had a reasonable expectation of success in using higher sampling frequencies (e.g., 48 kHz) in Chen's system because they reflect standard sampling frequencies widely known and used in the art. There are a finite number of sampling frequencies audio processing devices use. Moreover, downsampling audio signals to reduce computational requirements was well-known. Thus, implementing the sampling frequencies from Kim and Kleffner

(e.g., 48 kHz) in Chen’s system, and then downsampling for the noise reduction stage (e.g., to 8 kHz as taught by Li) would have been trivial.

167. Accordingly, for at least the reasons I discuss above, (i) Chen, Li, Kim, and Kleffner and/or (ii) Chen, Li, Florencio, Kim, and Kleffner disclose or render obvious each limitation of claim 10, and therefore render claim 10 obvious as a whole.

2. Claim 11

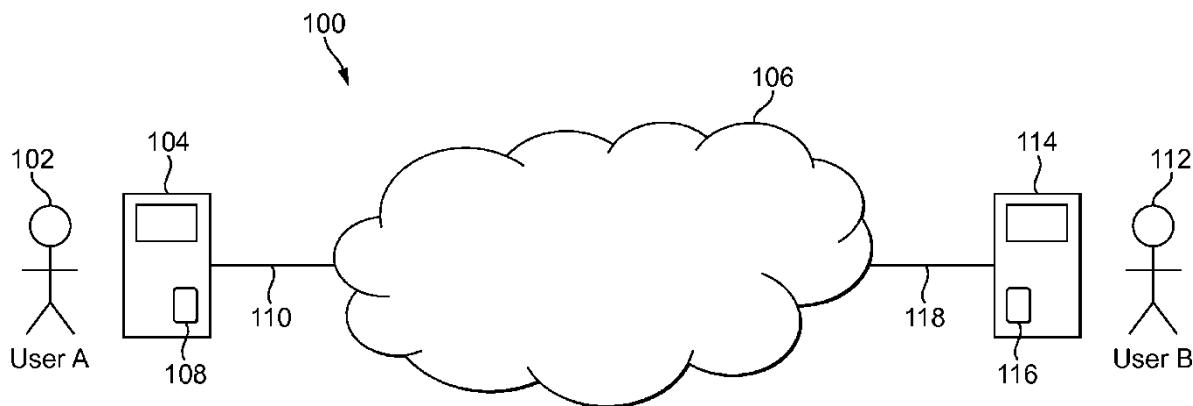
168. Claim 11 depends from claim 1 and further recites that “the speech segment determiner outputs more accurate speech segment information to the voice direction detector than speech segment information to the adaptive filter.” Kim and Kleffner disclose or render obvious this claim element.

169. As I discuss above for claim 10, a person of ordinary skill in the art would have been motivated to use an audio signal with a higher sampling frequency as taught by Kim in Chen’s (and Florencio’s) DOA unit and a lower sampling frequency in the noise reduction stage. (See ¶¶156-167, above.) A higher sampling frequency would result in more accurate speech segment information being provided to the voice direction detector (of Chen or Florencio), and a lower sampling frequency would result in less accurate speech information being provided to the adaptive filter (e.g., Li’s filter).

170. Accordingly, for at least the reasons I discuss above, (i) Chen, Li, Kim, and Kleffner and/or (ii) Chen, Li, Florencio, Kim, and Kleffner disclose or render obvious each limitation of claim 11, and therefore render claim 11 obvious as a whole.

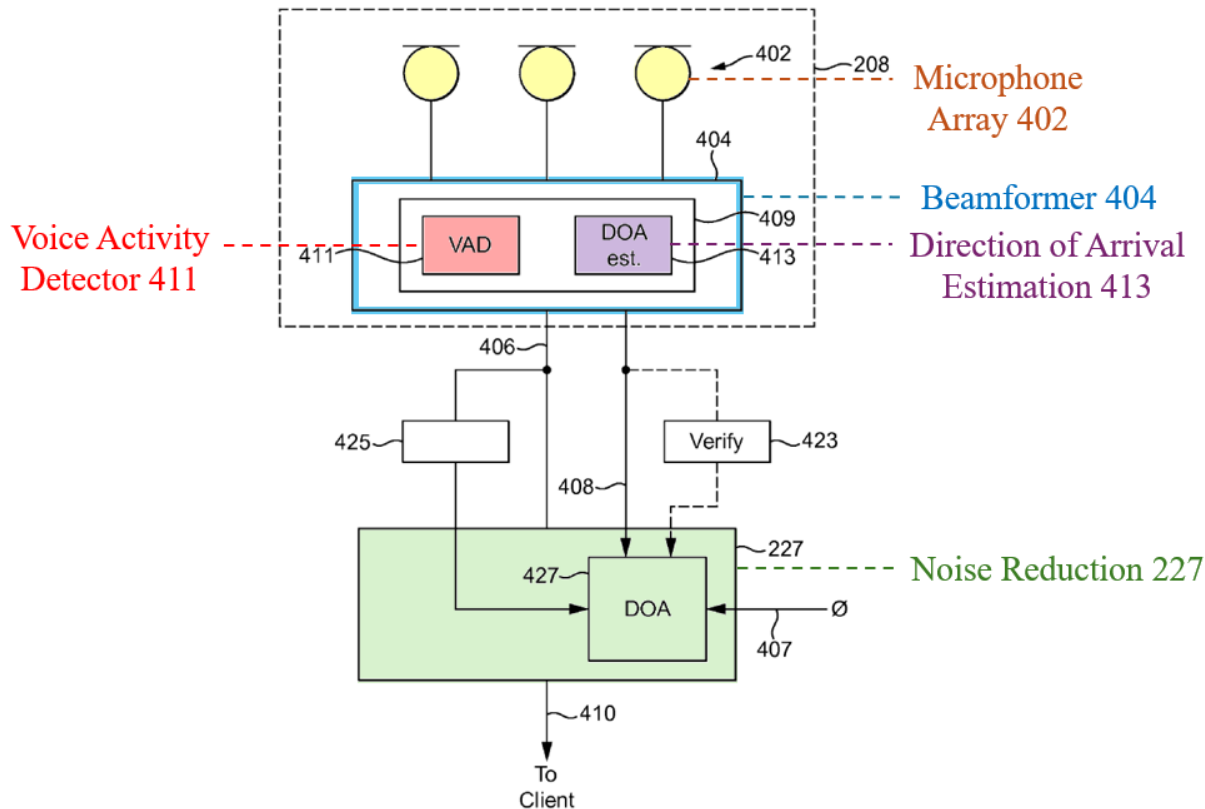
E. Claims 1-3, 9, and 13-15 Would Have Been Obvious in View of Strömmer.

171. Strömmer discloses a system for processing audio signals in a communication system. (EX-1010 (Strömmer), Abstract, ¶[0041].) The system includes a device (e.g., mobile phone or computer) that includes a microphone array connected to a beamformer. (*Id.* ¶¶[0041], [0053].) Strömmer depicts an example of a communication device in Figure 1, showing users (102, 112) and user devices (104, 114).



(*Id.*, Fig. 1.) The user devices communicate over the communications system via a network (106). (*Id.* ¶[0042].)

172. Strömmer's beamformer contains a voice activity detector (VAD), which determines if the audio signal contains speech by detecting “speech like qualities” in the signals. (EX-1010 (Strömmer) ¶¶[0041], [0053].) The beamformer also contains a direction of arrival (DOA) estimation unit. (*Id.*; *id.* ¶¶[0060]-[0063].) Using information from the VAD and DOA estimation block, the beamformer's processing block 409 determines the principal direction of the speaker. (*Id.* ¶[0053].) The beamformer 404 forms a beam with high gain in the direction of the wanted signal, e.g. speech, and a low gain in any other direction. (*Id.*) Strömmer's system, shown in Figure 4, also includes a subsequent noise reduction stage 227 that further reduces noise:



(*Id.*, Fig. 4, ¶¶[0056]-[0059].)

1. Claim 1

a. Preamble

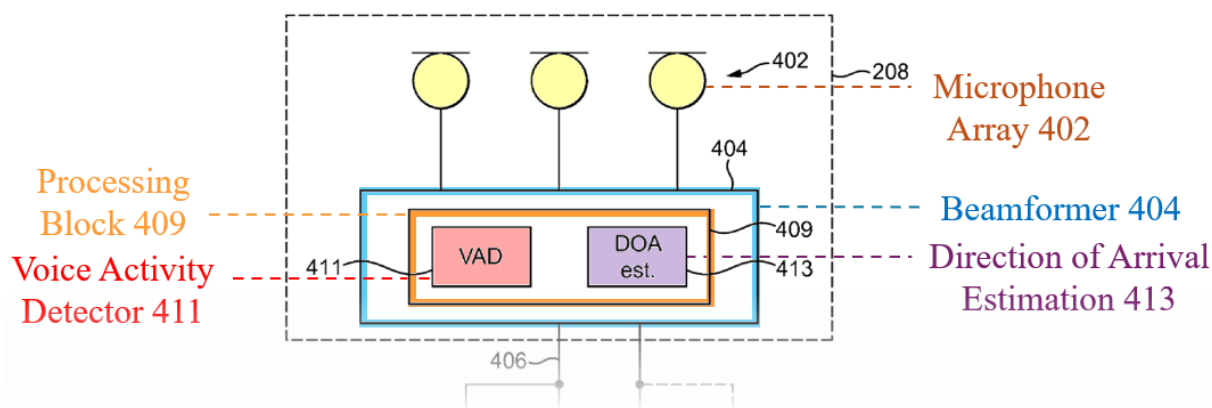
173. Strömmer discloses a “noise reduction apparatus” as recited in the preamble of claim 1 because it discloses a device that reduces noise in audio signals using a beamformer and a “noise reduction stage.” (EX-1010 (Strömmer), Abstract, ¶¶[0053] (beamforming to emphasize wanted speech signals and reduce noise), [0064]-[0065], [0057], Fig. 4.) Accordingly, Strömmer discloses the preamble of claim 1.

b. Element 1[a]: Speech Segment Determiner

174. Claim element 1[a] recites “a speech segment determiner configured to [i] determine whether or not a sound picked up by at least either a first microphone or a second microphone is a speech segment and [ii] to output speech segment information when it is determined that the sound picked up by the first or the second microphone is the speech segment.” Strömmer discloses or renders obvious this claim element.

175. Strömmer discloses a microphone array 402 and a beamformer 404. (EX-1010 (Strömmer) ¶¶[0042], [0052].) The output of each microphone “is coupled to the beamformer 404.” (*Id.* ¶[0052].) The beamformer contains a processing block 409 comprising a voice activity detector (VAD) 411. (*Id.* ¶[0053].) The VAD determines whether speech is present by detecting “speech like qualities.” (*Id.*;

id. ¶¶[0067], [0047] (identifying “desired audio signals . . . based on the detection of speech like qualities.”), claim 19 (detecting speech characteristics in the received signal).) Strömmer’s Figure 4 shows the microphone array, beamformer, and VAD:



(*Id.*, Fig. 4 (excerpt).) Strömmer’s microphone array could include two or more microphones. (*Id.* ¶[0052] (disclosure not limited to the three microphones shown in Figure 4).)

176. Accordingly, Strömmer discloses a speech segment determiner (e.g., VAD 411) that determines whether a sound picked up by at least either a first microphone or a second microphone (e.g., microphones in array 402) is a speech segment as recited in element 1[a][i].

177. Strömmer further discloses the VAD outputs speech segment information to processing block 409 and beamformer 404. For example, Strömmer generally describes how its system uses information from the VAD, and particularly

describes how the processing block uses such speech segment information from the VAD:

Audio signals that arrive at the microphone 208 from directions which have been identified as from a wanted source may be identified based on the detection of speech like characteristics and identified as being from a principal direction of a main speaker.

(EX-1010 (Strömmer) ¶[0067].)

Desired audio signals are identified when the audio signals are processed having been received at the microphone 208. During processing, desired audio signals are identified based on the detection of speech like qualities and a principal direction of a main speaker is determined.

(*Id.* ¶[0047].)

The processing block 409 ascertains the nature of the audio signals received by the microphone array 402, and based on detection of speech like qualities detected by the VAD 11 and DOA information estimated in block 413, one or more principal direction(s) of main speaker(s) is determined.

(*Id.* ¶[0053] (emphasis added.)) The beamformer 404 then forms a beam with a high gain in the direction of the wanted speaker signal and low gain in any other direction:

The beamformer 404 uses the DOA information to process the audio signals by forming a beam that has a high gain in the direction from the one or more principal direction(s) from which wanted signals are received at the microphone array and a low gain in any other direction.

(*Id.*)

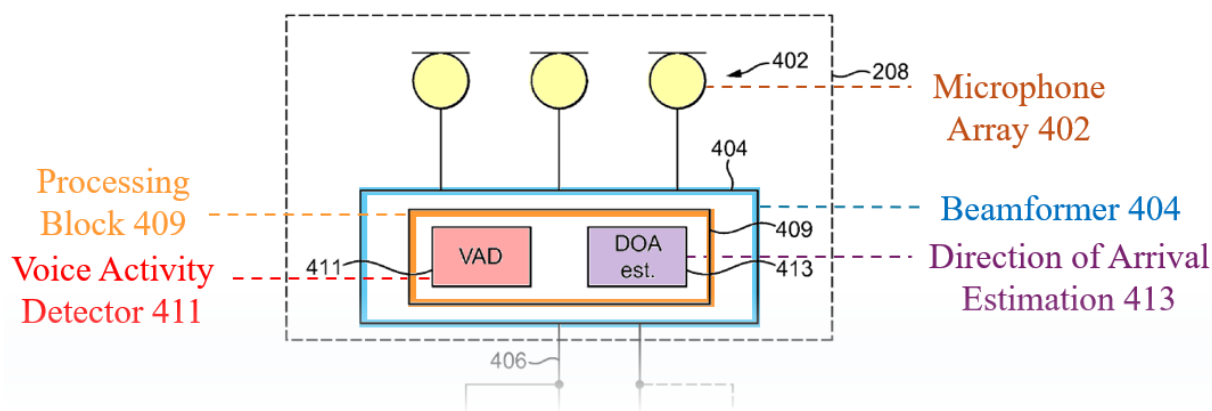
178. Accordingly, Strömmer discloses or renders obvious that the speech segment determiner (e.g., VAD) is configured to output (e.g., to processing block 409 and beamformer 404) speech segment information (e.g., that the audio signal

contains wanted signals, e.g., those with speech like qualities), when it is determined that the sound picked up by the first or the second microphone is the speech segment.

c. Element 1[b]: Voice Direction Detector

179. Element 1[b] recites “a voice direction detector configured, when receiving the speech segment information, [i] to detect a voice incoming direction indicating from which direction a voice sound travels, based on a first sound pick-up signal obtained based on a sound picked up by the first microphone and a second sound pick-up signal obtained based on a sound picked up by the second microphone and [ii] to output voice incoming-direction information when the voice incoming direction is detected.” Strömmer discloses or renders obvious this claim element.

180. In addition to the VAD 11, Strömmer’s processing block 409 contains a direction of arrival (DOA) estimation block 413:



(EX-1010 (Strömmer), Fig. 4 (excerpt).) The DOA estimation block 413 estimates DOA information by estimating the time delay between received audio signals at a

plurality of microphones, and estimating the source of the audio signal using a priori knowledge about the location of the plurality of microphones. (*Id.* ¶¶[0061], [0053], [0062] (“direction of arrival ... estimated” by block 413).) Strömmer discloses that:

In the DOA estimation block 413, the DOA information is estimated by estimating the time delay e.g. using correlation methods, between received audio signals at a plurality of microphones, and estimating the source of the audio signal using the a priori knowledge about the location of the plurality of microphones.

FIG. 5 shows microphones 403 and 405 receiving audio signals from an audio source 516. The direction of arrival of the audio signals at microphones 403 and 405 separated by a distance, d can be estimated using equation (1):

$$\theta = \arcsin\left(\frac{\tau_D v}{d}\right) \quad (1)$$

where v is the speed of sound, and τ_D is the difference between the times the audio signals from the source 516 arrive at the microphones 403 and 405—that is, the time delay. The time delay is obtained as the time lag that maximises the cross-correlation between the signals at the outputs of the microphones 403 and 405. The angle θ may then be found which corresponds to this time delay.

It will be appreciated that calculating a cross-correlation of signals is a common technique in the art of signal processing and will not be describe in more detail herein.

(*Id.* ¶¶[0061]-[0063].)

181. Based on speech detection by the VAD 11 and DOA information estimated in block 413, processing block 409 determines the “principal direction(s) of main speaker(s)[.]” (EX-1010 (Strömmer) ¶[0053]; *id.* (processing block 409 determines principal direction), ¶¶[0047] (“desired audio signals are identified based on

the detection of speech like qualities and a principal direction of a main speaker is determined”), [0067].)

182. Accordingly, Strömmer discloses or renders obvious a voice direction detector (e.g., processing block 409) configured, when receiving the speech information (e.g., from VAD 11), to detect a voice incoming direction indicating from which direction a voice sound travels (e.g., the “principal direction(s) of main speaker(s)”), based on a first sound pick-up signal obtained based on a sound picked up by the first microphone and a second sound pick-up signal obtained based on a sound picked up by the second microphone (e.g., the signals obtained by the first and second microphones in array 402) as recited in element 1[b][i].

183. Strömmer also discloses that the voice direction detector (e.g., processing block 409) is configured to output voice incoming-direction information as recited in element 1[b][ii]. Strömmer discloses outputting the voice incoming-direction information in two ways. First, Strömmer’s voice processing block 409 outputs the “DOA information” to beamformer 404, which “uses the DOA information to process the audio signal by forming a beam.” (EX-1010 (Strömmer) ¶[0053]; *id.* ¶[0067].) Second, Strömmer discloses that “DOA information estimated in the beamformer 404 is supplied to the noise reduction stage 227 and to signal processing circuitry 420.” (*Id.* ¶[0055].)

184. Accordingly, Strömmer discloses outputting (e.g., to beamformer 404 or the noise reduction stage) the voice incoming-direction information (e.g., DOA information) when the voice incoming direction is detected. Thus, Strömmer discloses or renders obvious element 1[b][ii].

d. Element 1[c]: Adaptive Filter

185. Element 1[c] recites “an adaptive filter configured to perform a noise reduction process using the first and second sound pick-up signals based on the speech segment information and the voice incoming-direction information.” Strömmer discloses or renders obvious this claim element. Strömmer discloses two adaptive filters: (1) beamformer 404 and (2) noise reduction stage 227.

i. Beamformer

186. A beamformer is a spatial filter that reduces noise by suppressing the signal from all directions except the desired ones. (EX-1011 (Brandstein), 3.) An “adaptive” beamformer is an “adaptive filter.” Indeed, Strömmer teaches that adaptive beamformers “use the DOA information to filter the signals from the microphones in an array[.]” (EX-1010 (Strömmer) ¶[0008].)

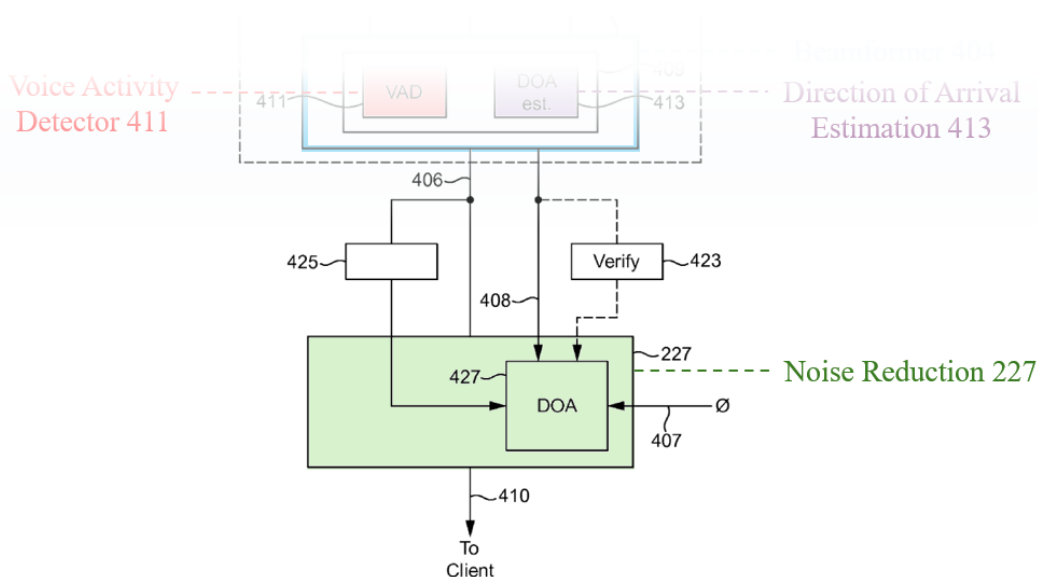
187. Strömmer discloses an adaptive beamformer 404 that “uses the DOA information to process the audio signals by forming a beam that has a high gain in the direction from the one or more principal direction(s) from which wanted signals are received at the microphone array and a low gain in any other direction.” (EX-

1010 (Strömmer) ¶¶[0053].) As I discuss above, Strömmer's adaptive beamformer uses (via processing block 409) information from the VAD regarding whether the audio signal contains speech and DOA information from DOA estimator 413 to determine the direction of arrival for wanted signals (e.g. speech). (See ¶¶174-184, above.)

188. Accordingly, Strömmer discloses or renders obvious an adaptive filter (e.g., adaptive beamformer) configured to perform a noise reduction process using the first and second sound pick-up signals (e.g., the signals obtained by the first and second microphones in array 402) based on the speech segment information (e.g., from VAD 11) and the voice incoming-direction information (e.g., DOA information including the direction of the speaker).

ii. Noise Reduction

189. Strömmer also discloses a noise reduction stage 227:



(EX-1010 (Strömmer), Fig. 4 (excerpt); *id.* ¶¶[0053], [0055]-[0059], [0064]-[0071].)

The noise reduction stage receives DOA information and audio output from the beamformer. (*Id.* ¶¶[0053], 0055], [0064], [0068] (DOA information includes the angle from which a main component of the audio was received at the microphone).) The beamformer audio output contains a filtered combination of the microphone pickup signals. (*Id.* ¶[0008]; EX-1016 (Benesty), 946-953.)

190. The noise reduction stage compares the DOA information provided from the beamformer (e.g., direction of main component) to DOA information 427 known to the terminal (e.g., direction of wanted source), and processes the beamformer output based on that comparison. (EX-1010 (Strömmer) ¶[0068].) For example, if the DOA information from the beamformer indicates the main component in the audio signal comes from a wanted direction, then the noise reduction stage “determines a level of noise suppression using conventional methods[.]” (*Id.* ¶[0071].) But if the DOA information indicates the main component comes from an unwanted direction, the signals are treated as noise and the noise reduction unit will apply “maximum attenuation.” (*Id.* ¶[0072].) The noise reduction stage is an adaptive filter because it adaptively changes the attenuation level based on current DOA information.

191. Accordingly, Strömmer discloses or renders obvious an adaptive filter (e.g., noise reduction stage) configured to perform a noise reduction process using the first and second sound pick-up signals (e.g., output of the beamformer, which

contains a filtered combination of the signals from the first and second microphones) based on the speech segment information (e.g., whether the signal contains speech-like qualities) and the voice incoming-direction information (e.g., DOA information).

192. Accordingly, for at least the reasons I discuss above, Strömmer discloses or renders obvious each limitation of claim 1, and therefore renders claim 1 obvious as a whole.

2. Claim 13

193. As I discuss above, claim 13 recites a method claim for performing the same limitations as the apparatus of claim 1. (See ¶68, above.) Accordingly, Strömmer renders obvious claim 13 for the reasons I discuss above for claim 1. (See ¶¶173-192, above.)

3. Claims 2 and 14

194. Claim 2 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on a phase difference between the first and second sound pick-up signals.” Claim 14 depends from claim 13 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals.” Strömmer discloses or renders obvious these claim elements.

195. Strömmer discloses using the time difference estimation (e.g., phase difference) to determine the direction of arrival. (EX-1010 (Strömmer) ¶[0061]; *see also* EX-1001 ('259 patent), 18:41-47 (equating phase difference with time delay).) Time delay is the “difference between the times the audio signals from the source 516 arrive at the microphones 403 and 405.” (EX-1010 (Strömmer) ¶[0062].) Strömmer illustrates this delay in Figure 5, which shows two microphones 403 and 405 receiving audio signals from a source 516:

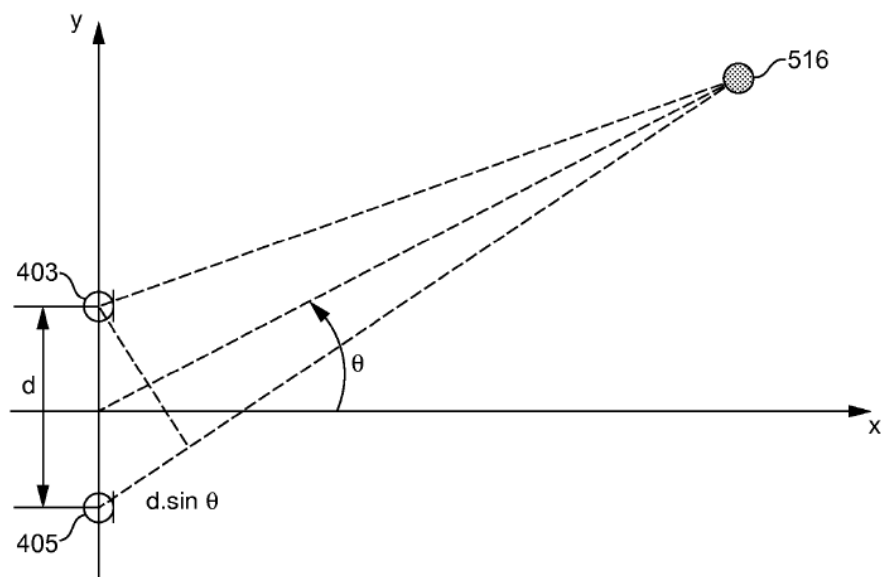


FIG. 5

(*Id.*, Fig. 5.)

196. Strömmer explains that its system estimates the time delay “using correlation methods, between received audio signals at a plurality of microphones, and estimating the source of the audio signal using the a priori knowledge about the location of the plurality of microphones.” (EX-1010 (Strömmer) ¶[0061].) “The time

delay is obtained as the time lag that maximises the cross-correlation between the signals at the outputs of the microphones 403 and 405.” (*Id.* ¶[0062].) Strömmer explains that “calculating a cross-correlation of signals is a common technique in the art of signal processing.” (*Id.* ¶[0063].) A person of ordinary skill in the art would have understood Strömmer’s cross-correlation time delay estimation uses a phase difference between the two detected audio signals to determine the sound source direction. (EX-1016 (Benesty), 1046-1047.)

197. Thus, Strömmer discloses or renders obvious detecting the voice incoming direction based on a phase difference (e.g., time delay) between the first and second sound signals.

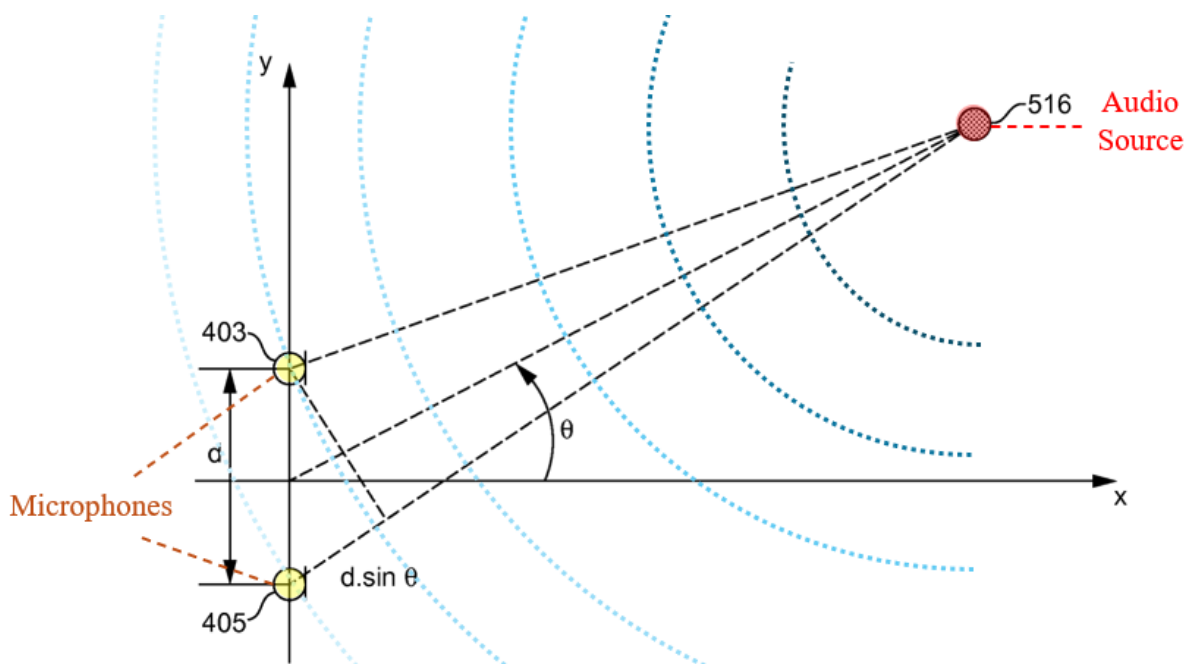
198. Accordingly, for at least the reasons I discuss above, Strömmer discloses or renders obvious each limitation of claims 2 and 14, and therefore renders claims 2 and 14 obvious as a whole.

4. Claims 3 and 15

199. Claim 3 depends from claim 2 and further recites that “the adaptive filter performs the noise reduction process to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first sound pick-up signal has a more advanced phase than the second sound pick-up signal whereas the adaptive filter performs the noise reduction process to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up

signal when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal.” Claim 15 depends from claim 14 and further recites that “the noise reduction process is performed to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first sound pick-up signal has a more advanced phase than the second sound pick-up signal whereas the noise reduction process is performed to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up signal when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal.” Strömmer renders obvious these claim elements.

200. As I discuss above for claims 2 and 14, Strömmer discloses using the phase difference (e.g., time delay) to determine the DOA. (See ¶¶194-198, above.) Strömmer illustrates determining the direction of the audio source in Figure 5:



(EX-1010 (Strömmer), Fig. 5.) As shown by the dotted blue lines, which are added

to the Figure for illustration, the sound waves from the audio source arrive at the microphones (403, 405) at different times. A person of ordinary skill in the art would have understood that the microphone that detects the sound from the audio source first will be closest to the audio source. In other words, the sound signal detected by the first microphone will have a more advanced phase if the sound source is closer to that first microphone. In Figure 5, the audio signal detected by microphone 403 has a more advanced phase than the audio signal detected by microphone 405. If the audio source were below the x-axis in Figure 5, the signal detected by microphone 405 would have a more advanced phase than the audio signal detected by microphone 403.

201. Strömmer's beamformer 404 forms a beam with "a high gain in the direction" of the speaker and "a low gain in any other direction." (EX-1010 (Strömmer) ¶[0053].) To form the beam, Strömmer applies coefficients to the microphone array signals. (See EX-1011 (Brandstein), 8-11.) This beam constructively adds the signals from the speaker (high gain) and destructively adds signals from other directions (low gain). (*Id.*) Thus, a noise component in the first signal (e.g., sound detected by microphone 403) is reduced using the second signal (e.g., sound detected by microphone 405) because the noise from the second signal is destructively added to the noise in the first signal to reduce it.

202. Thus, a person of ordinary skill in the art would have understood that Strömmer reduces a noise component carried by the first sound pick-up signal (e.g., sound detected by microphone 403) using the second sound pick-up signal (e.g., sound detected by microphone 405) when the first sound pick-up signal has a more advanced phase than the second sound pick-up signal; and reduces a noise component carried by the second sound pick-up signal (e.g., sound detected by microphone 405) using the first sound pick-up signal (e.g., sound detected by microphone 403) when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal.

203. Accordingly, for at least the reasons I discuss above, Strömmer discloses or renders obvious each limitation of claims 3 and 15, and therefore renders claims 3 and 15 obvious as a whole.

5. Claim 9

204. Claim 9 depends from claim 1 and further recites that “the speech segment determiner detects the speech segment based on the first sound pick-up signal when the first sound pick-up signal has a more advanced phase than the second sound pick-up signal whereas the speech segment determiner detects the speech segment based on the second sound pick-up signal when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal.” Strömmer renders obvious this claim element.

205. Strömmer's VAD detects speech in each of the received sound signals. (EX-1010 (Strömmer) ¶[0053].) Thus, a person of ordinary skill in the art would have understood that Strömmer's VAD detects speech "based on" the first sound pick-up signal (e.g., signal from first microphone) when the first sound pick-up signal has a more advance phase (e.g., speaker is closer to first microphone) and detects speech "based on" the second sound pick-up signal (e.g., signal from second microphone) when the second sound pick-up signal has a more advanced phase (e.g., speaker is closer to second microphone).

206. Accordingly, for at least the reasons I discuss above, Strömmer discloses or renders obvious each limitation of claim 9, and therefore renders claim 9 obvious as a whole.

F. Claims 4, 12, and 16 Would Have Been Obvious in View of Strömmer and Visser.

207. Claims 4 and 16 depend from claims 2 and 14, respectively. Furthermore, claim 12 includes limitations that mirror those of claim 1. As I discuss above, Strömmer discloses or renders obvious each limitation of claims 1, 2, and 14. (See ¶¶173-198, above.) For at least those reasons and the reasons I discuss below, claims 4, 12, and 16 would have been obvious in further view of Visser. I incorporate my above discussion of Strömmer here, and only discuss claim elements for which Visser is relevant in combination with Strömmer.

1. Claims 4 and 16

208. Claim 4 depends from claim 2 and further recites that “when the phase difference is within a predetermined range, the adaptive filter outputs either the first or the second sound pick-up signal without performing the noise reduction process.” Claim 16 depends from claim 14 and further recites the same limitation. As I discuss above, Visser discloses or renders obvious these claim elements. (See ¶¶124-131, above.)

209. A person of ordinary skill in the art would have been motivated to incorporate Visser’s processing of broadside signals into Strömmer’s device, and would have reasonably expected success in doing so, for the same reasons a person of ordinary skill would have been motivated to do so for Chen. (See ¶¶124-131, above.)

210. Accordingly, for at least the reasons I discuss above, Strömmer and Visser disclose or render obvious each limitation of claims 4 and 16, and therefore render claims 4 and 16 obvious as a whole.

2. Claim 12

211. As I discuss above (see ¶¶100-101, above), claim 12 recites two claim elements in addition to those recited in claim 1: element 12[a] recites “a first face and an opposite second face that is apart from the first face with a specific distance” and element 12[b] recites “a first microphone and a second microphone provided on

the first face and the second face, respectively.” As I discuss above, Visser discloses these claim elements. (*See* ¶¶146-152, above.)

212. A person of ordinary skill in the art would have been motivated to incorporate Visser’s microphone arrangement into Strömmer’s mobile device, and would have reasonably expected success in doing so, for the same reasons a person of ordinary skill would have been motivated to do so for Chen, as I discuss above. (*See* ¶¶146-152, above.)

213. Accordingly, for at least the reasons I discuss above, Strömmer and Visser disclose or render obvious each limitation of claim 12, and therefore render claim 12 obvious as a whole.

G. Claims 5-6 and 17-18 Would Have Been Obvious in View of Strömmer and Brandstein.

214. Claims 5-6 and 17-18 depend, directly or indirectly, from claims 1 and 14. As I discuss above, Strömmer discloses or renders obvious each limitation of claims 1 and 14. (*See* ¶¶173-198, above.) For at least those reasons and the reasons I discuss below, claims 5-6 and 17-18 would have been obvious in further view of Brandstein. I incorporate my above discussion of Strömmer here.

215. Brandstein is a textbook on digital signal processing of microphone arrays. (EX-1011.) I purchased a copy of the Brandstein textbook sometime between 2001 and 2003. Exhibit 1011 is a scan of my copy of Brandstein, as evidenced by my name written on one of the first pages. Brandstein discloses multiple techniques

for SSL and speech recognition. (*E.g., id.*, 157-201 (SSL), 331-354 (speech recognition).)

1. Claims 5 and 17

216. Claim 5 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on magnitudes of the first and second sound pick-up signals.” Claim 17 depends from claim 14 and further recites that “the voice incoming direction is detected based on magnitudes of the first and second sound pick-up signals.” Brandstein discloses these claim elements.

217. Brandstein describes three general categories of SSL: (1) maximizing the steered response power (SRP), (2) high-resolution spectral estimation, and (3) time-difference of arrival. (EX-1011 (Brandstein), 158.)

218. In SRP-based SSL, a beamformer is steered “to specific spatial points of interest in some fashion” to evaluate the output signal, typically the power. (EX-1011 (Brandstein), 170.) “When the focus corresponds to the location of the sound source, the SRP should reach a global maximum.” (*Id.*) The SRP is a function of the beamformer output’s magnitude, as shown in Brandstein’s Equation 8.16:

$$P(\mathbf{q}) = \int_{-\infty}^{+\infty} |Y(\omega)|^2 d\omega$$

(*Id.*) And the beamformer output is a function of the microphone signals, as shown in Equation 8.15:

$$Y(\omega, q) = \sum_{n=1}^N G_n(\omega) X_n(\omega) e^{j\omega \Delta_n}$$

(*Id.*) A person of ordinary skill in the art would have understood based on these equations that the SRP is a function of the microphone signals' magnitudes.

219. Brandstein also discloses SRP SSL in the frequency domain using a phase transform (SRP-PHAT). (EX-1011 (Brandstein), 170-172.) SRP-PHAT is “similar to the standard SRP-based approaches.” (*Id.*, 171.) Specifically, SRP-PHAT looks for maximum power over a region of potential source locations. (*Id.*, 172.) But SRP-PHAT includes a “PHAT weighting” that “has been found to provide an enhanced robustness in low to moderate reverberation conditions.” (*Id.*, 170.) This PHAT weighting is a function of the magnitude of each microphone signal, as highlighted below in Brandstein's Equation 8.20:

$$G_n(\omega) = \frac{1}{|X_n(\omega)|}$$

(*Id.*, 171.) Thus, Brandstein discloses two methods of determining the direction of a sound source using microphone signal magnitudes—SRP and SRP-PHAT. (*See* EX-1012 (Visser), 28:60-64 (SRP-PHAT is based on gain differences).)

220. Accordingly, Brandstein discloses a voice direction detector that detects the voice incoming direction based on magnitudes of the first and second sound pick-up signals (e.g., SRP and SRP-PHAT).

221. A person of ordinary skill in the art would have been motivated to use SRP or SRP-PHAT SSL in Strömmer's system for several reasons.

222. First, Brandstein explains that SRP and SRP-PHAT offer advantages in certain environments. With respect to SRP, its simple beamformer is "optimal" for environments that have a "limited case of additive, uncorrelated noise, and uniform variance noise and equal source-microphone distances." (EX-1011 (Brandstein), 169.) Thus, a person of ordinary skill would have been motivated to use SRP in Strömmer's system when the environment of the system meets these criteria.

223. Second, SRP-PHAT seeks to combine the "advantages of the steered beamformer for source localization with the signal and condition independent robustness offered by the PHAT weighting." (EX-1011 (Brandstein), 171.) SRP-PHAT's frequency domain analysis provides an "enhanced robustness in low to moderate reverberation conditions." (*Id.*, 170.) SRP-PHAT also "requires shorter analysis intervals and exhibits an elevated insensitivity to environmental conditions." (*Id.*, 171.) SRP-PHAT also has "decreased sensitivity to noise and reverberations and [provides] more precise location estimates than the existing location methods offer." (*Id.*, 172.) These advantages would have motivated a person of ordinary skill to replace the cross-correlation SSL in Strömmer with the SRP-PHAT SSL in Brandstein. Indeed, Brandstein describes experiments showing SRP-PHAT outperformed "TDOA-based localization methods," which are the type of methods

that Strömmer uses. (*Id.*, 177; *id.*, 161 (TDOA used cross-correlation); EX-1010 (Strömmer) ¶¶[0061]-[0063] (relying on cross-correlation).)

224. Third, SRP and SRP-PHAT were known SSL techniques (EX-1011 (Brandstein), 158) and using them in Strömmer would have yielded predictable results (e.g., more accurate SSL). Fourth, because SRP and SPR-PHAT were standard SSL techniques, using SRP or SRP-PHAT in Strömmer's system would have been a simple substitution of one known SSL technique (e.g., SRP or SRP-PHAT) for another (e.g. Strömmer's cross-correlation) to obtain predictable results (e.g., accurate SSL).

225. Fifth, a person of ordinary skill would have found it obvious to try other SSL techniques like SRP and SRP-PHAT because there were only a finite number of known SSL techniques. (*See* EX-1011 (Brandstein), 158.)

226. A person of ordinary skill would have had a reasonable expectation of success in using SRP or SPR-PHAT techniques in Strömmer's system. Brandstein describes multiple SSL methods, including the cross-correlation methods as in Strömmer. (EX-1011 (Brandstein), 161-162.) Thus, a person of ordinary skill would have reasonably expected that implementing the well-known SRP or SPR-PHAT techniques from Brandstein in Strömmer's system would work with, and improve, Strömmer's noise reduction system.

227. Accordingly, for at least the reasons I discuss above, Strömmer and Brandstein disclose or render obvious each limitation of claims 5 and 17, and therefore render claims 5 and 17 obvious as a whole.

2. Claims 6 and 18

228. Claim 6 depends from claim 5 and further recites that “the adaptive filter performs the noise reduction process to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first sound pick-up signal has a greater magnitude than the second sound pick-up signal whereas the adaptive filter performs the noise reduction process to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up signal when the second sound pick-up signal has a greater magnitude than the first sound pick-up signal.” Claim 18 depends from claim 17 and further recites that “the noise reduction process is performed to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first sound pick-up signal has a greater magnitude than the second sound pick-up signal whereas the noise reduction process is performed to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up signal when the second sound pick-up signal has a greater magnitude than the first sound pick-up signal.” Strömmer and Brandstein render obvious these claim elements.

229. Strömmer's beamformer 404 forms a beam with "a high gain in the direction" of the speaker and "a low gain in any other direction." (EX-1010 (Strömmer) ¶[0053].) To form this beam, signals from Strömmer's first and second microphones (403, 405) are filtered such that speech components (from a desired direction) are emphasized while noise components (from undesired directions) cancel each other out. (See EX-1011 (Brandstein), 8-11, 88-90.)

230. The cancellation of noise components in a beamformer is mutual. Noise components in a first signal cancel the noise components in a second signal. At the same time, noise components in the second signal cancel the noise components in the first signal. This mutual cancellation of noise components occurs regardless of which microphone signal has a greater magnitude.

231. As I discuss above, Strömmer's beamformer reduces a noise component carried by the first sound pick-up signal (e.g., sound detected by microphone 403) using the second sound pick-up signal (e.g., sound detected by microphone 405) when the first sound pick-up signal has a greater magnitude than the second sound pick-up signal; and reduces a noise component carried by the second sound pick-up signal (e.g., sound detected by microphone 405) using the first sound pick-up signal (e.g., sound detected by microphone 403) when the second sound pick-up signal has a greater magnitude than the first sound pick-up signal. (See ¶¶199-203, 229-230, above.)

232. Accordingly, for at least the reasons I discuss above, Strömmer and Brandstein disclose or render obvious each limitation of claims 6 and 18, and therefore render claims 6 and 18 obvious as a whole.

H. Claims 7-8 and 19-20 Would Have Been Obvious in View of Strömmer, Brandstein, and Visser.

233. Claims 7-8 and 19-20 depend, directly or indirectly, from claims 1, 5, and 14. As I discuss above, Strömmer and/or Brandstein disclose or render obvious each limitation of claims 1, 5, and 14. (*See* ¶¶173-198, 216-227, above.) For at least those reasons and the reasons I discuss below, claims 7-8 and 19-20 would have been obvious in further view of Visser. I incorporate my above discussion of Strömmer and Brandstein here.

1. Claim 7

234. Claim 7 depends from claim 5 and further recites that “when a power difference that is a difference between magnitudes of the first and second sound pick-up signals is within a predetermined range, the adaptive filter outputs either the first or the second sound pick-up signal without performing the noise reduction process.” As I discuss above, Visser discloses or renders obvious the additional claim element of claim 7. (*See* ¶¶132-135, above.)

235. A person of ordinary skill in the art would have been motivated to incorporate Visser’s SSL and noise reduction for broadside sources into Strömmer’s device, and would have had a reasonable expectation of success in doing so, for the

same reasons as for incorporating the SSL and noise reduction into Chen. (*See* ¶¶124-131, above.)

236. Accordingly, for at least the reasons I discuss above, Strömmer, Brandstein, and Visser disclose or render obvious each limitation of claim 7, and therefore render claim 7 obvious as a whole.

2. Claims 8 and 19

237. Claim 8 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” Claim 19 depends from claim 14 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” Thus, claims 8 and 19 recite that voice incoming direction is based on a phase difference (as in claims 2 and 14) and magnitude (as in claims 5 and 17). As I discuss above, Visser discloses these claim limitations. (*See* ¶¶136-145, above.)

238. As I also discuss above, Strömmer discloses using phase difference, and Brandstein discloses using magnitude. (*See* ¶¶194-198, 216-227, above.) Thus, Strömmer, Brandstein, and Visser together render obvious detecting the voice in-

coming direction (e.g., DOA) based on a phase difference between the first and second sound pick-up signals (e.g., Strömmer's cross-correlation) and magnitudes of the first and second sound pick-up signals (e.g., Brandstein's SRP or SRP-PHAT).

239. A person of ordinary skill in the art would have been motivated to combine multiple SSL techniques in Strömmer's device, as taught by Visser, and would have had a reasonable expectation of success in doing so, for the same reasons I discuss above with Chen. (*See* ¶¶136-145, above.)

240. Accordingly, for at least the reasons I discuss above, Strömmer, Brandstein, and Visser disclose or render obvious each limitation of claims 8 and 19, and therefore render claims 8 and 19 obvious as a whole.

3. Claim 20

241. Claim 20 depends from claim 19 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” I note that the additional claim element of claim 20 is equivalent to the additional claim element of claim 9. (*See* ¶96, above.) As I discuss above (*see* ¶¶204-206, above), Strömmer renders obvious the additional claim element of claim 20 for the same reasons it renders obvious the additional claim element of claim 9.

242. Accordingly, for at least the reasons I discuss above, Strömmer, Brandstein, and Visser disclose or render obvious each limitation of claim 20, and therefore render claim 20 obvious as a whole.

I. Claims 10 and 11 Would Have Been Obvious in View of Strömmer, Kim, and Kleffner.

243. Claims 10 and 11 depend from claim 1. As I discuss above, Strömmer discloses or renders obvious each limitation of claim 1. (See ¶¶173-192, above.) For at least those reasons and the reasons I discuss below, claims 10 and 11 would have been obvious in further view of Kim and Kleffner. I incorporate my above discussion of Strömmer here.

1. Claim 10

244. Claim 10 depends from claim 1 and further recites that “signals are supplied to the voice direction detector as the first and second sound pick-up signals at a sampling frequency of 24 KHz or higher and signals are supplied to the adaptive filter as the first and second sound pick-up signals at a sampling frequency of 12 KHz or lower.” As I discuss above, Kim and Kleffner disclose or render obvious this claim element. (See ¶¶156-167, above.)

245. A person of ordinary skill in the art would have been motivated to combine the teachings of Kim and Kleffner, and to use the recited frequencies and downsampling in Strömmer, for several reasons.

246. First, Strömmer does not describe what sampling frequency to use. Thus, a person of ordinary skill would have been motivated to look to references such as Kim and Kleffner to see what sampling frequency to use.

247. Furthermore, a person of ordinary skill would have been motivated to incorporate Kim and Kleffner's sampling frequencies and downsampling into Strömmer's device, and would have reasonably expected success in doing so, for the same reasons I discuss above regarding Chen. (See ¶¶156-167, above.)

248. Accordingly, for at least the reasons I discuss above, Strömmer, Kim, and Kleffner disclose or render obvious each limitation of claim 10, and therefore render claim 10 obvious as a whole.

2. Claim 11

249. Claim 11 depends from claim 1 and further recites that "the speech segment determiner outputs more accurate speech segment information to the voice direction detector than speech segment information to the adaptive filter." Strömmer, Kim, and Kleffner render obvious this claim element.

250. As I discuss above for claim 10, a person of ordinary skill in the art would have been motivated to use an audio signal with a higher sampling frequency in a DOA unit (like Chen's or Strömmer's). (See ¶¶156-167, above.) A person of ordinary skill in the art would have understood that, like the DOA unit, Strömmer's

VAD (speech segment determiner) would provide better results if there is more information (e.g., higher sampling frequency). Thus, a person of ordinary skill would have been motivated to use a higher sampling frequency for Strömmer's VAD, and downsample after the VAD and DOA processes are completed to reduce the computational requirements of the beamforming and noise reduction units, where high sampling frequencies are unnecessary. Thus, for the same reasons I discuss above for claim 10, a person of ordinary skill would have been motivated to have Strömmer's VAD output more accurate speech segment information to the DOA unit (e.g., audio signal with a higher sampling frequency) and less accurate speech segment information (e.g., audio signal with a lower sampling frequency) to the adaptive filter (e.g., beamformer or noise reduction stage).

251. Accordingly, for at least the reasons I discuss above, Strömmer, Kim, and Kleffner disclose or render obvious each limitation of claim 11, and therefore render claim 11 obvious as a whole.

J. Claims 1-3, 9, and 13-15 Would Have Been Obvious in View of Florencio.

252. As I discuss above, Florencio's system includes a microphone array, a converter, a voice activity detector (306), an SSL unit (308), and an adaptive beamformer (310):

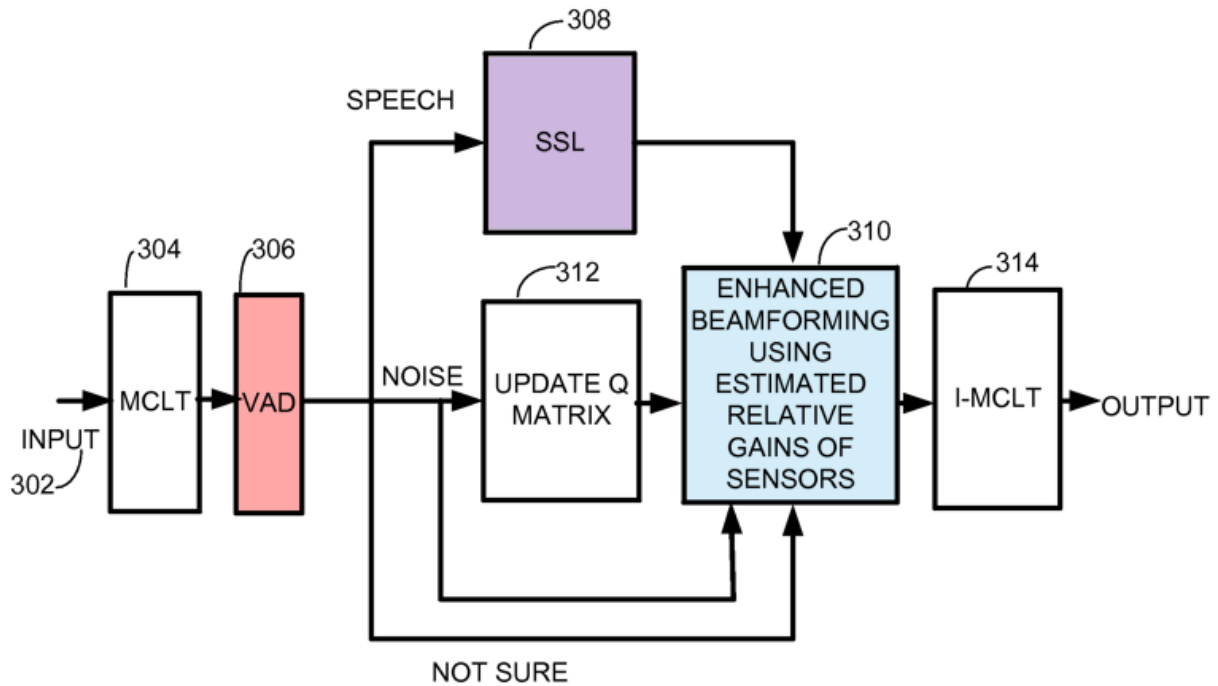


FIG. 3

(EX-1005 (Florencio), Fig. 3, ¶¶[0020]-[0024], [0044]-[0045], [0048], Figs. 1-7; *see* ¶105, above.) Florencio's process is shown in Figure 7:

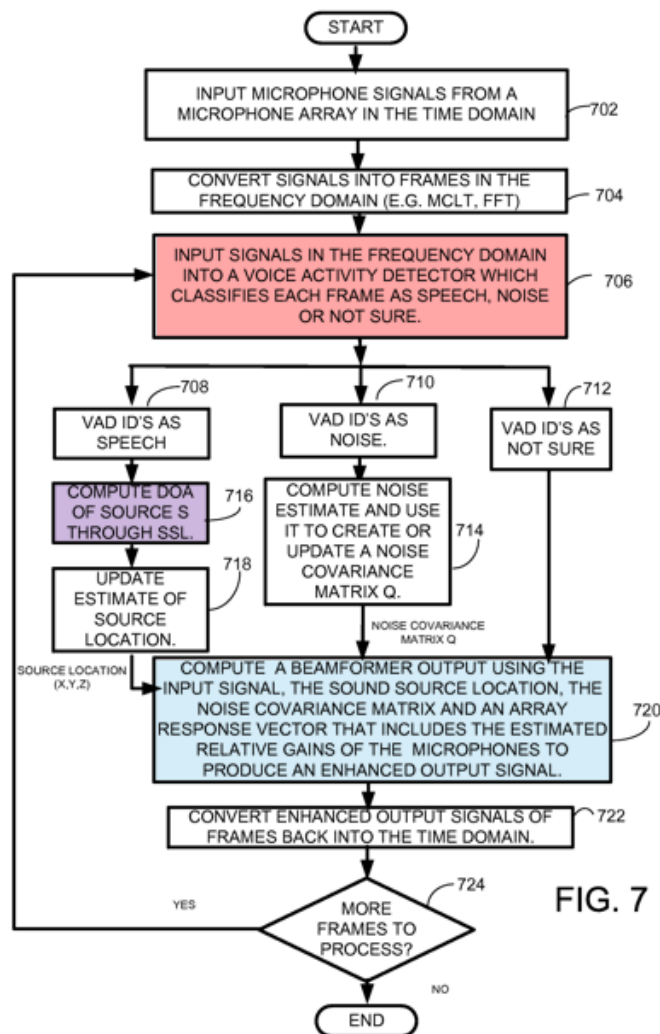


FIG. 7

(EX-1005 (Florencio), Fig. 7.)

1. Claim 1

a. Preamble

253. Florencio discloses a “noise reduction apparatus” as recited in the preamble because it discloses noise reduction using an “enhanced beamforming technique” that can be implemented on “a general purpose device.” (EX-1005 (Florencio) ¶¶[0012], [0005], [0002], Fig. 1.) An exemplary apparatus is shown in Figure 1:

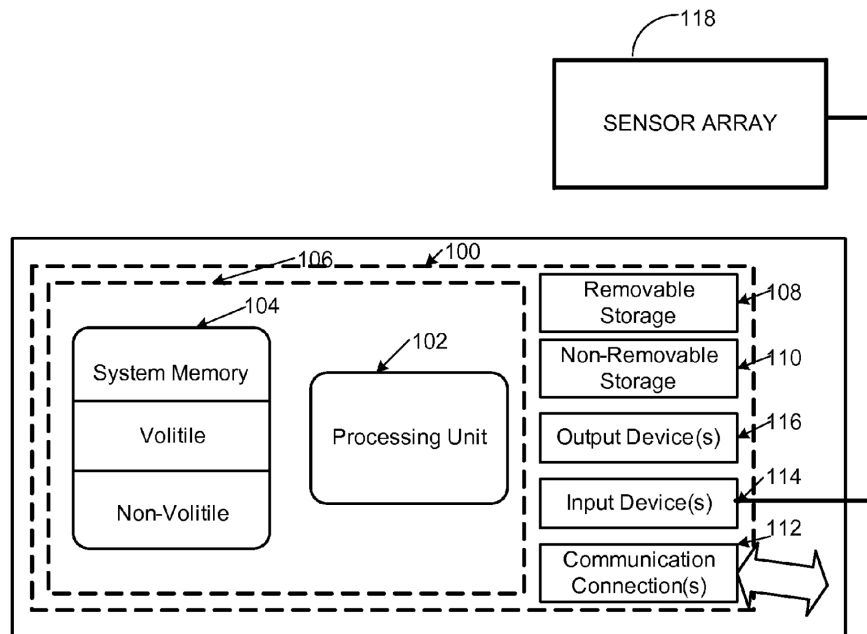


FIG. 1

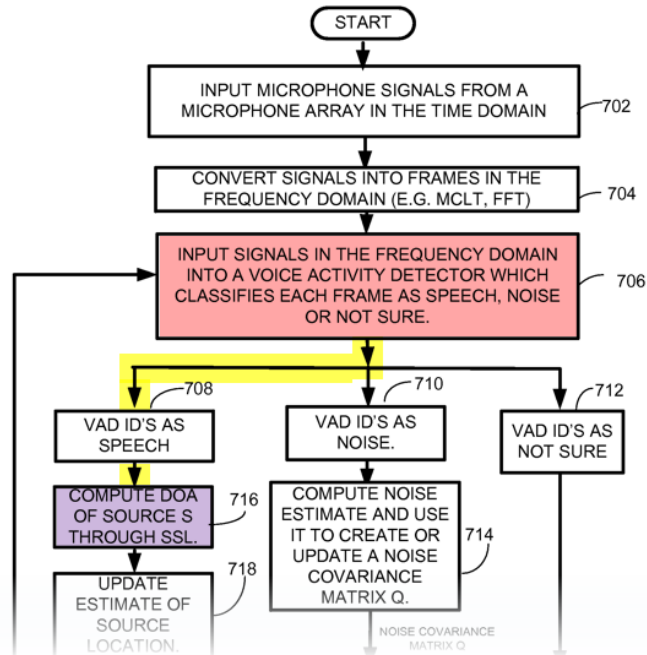
(*Id.*, Fig. 1.) Accordingly, Florencio discloses the preamble of claim 1.

b. Element 1[a]: Speech Segment Determiner

254. Claim element 1[a] recites “a speech segment determiner configured to [i] determine whether or not a sound picked up by at least either a first microphone or a second microphone is a speech segment and [ii] to output speech segment information when it is determined that the sound picked up by the first or the second microphone is the speech segment.” Florencio discloses or renders obvious this claim element.

255. Florencio receives signals at a microphone array and inputs the signals “into a Voice Activity Detector [which] classifies each input frame as Speech, Noise, or Not Sure.” (EX-1005 (Florencio) ¶¶[0044]-[0045], [0048], Figs. 3-4, 7.) If the

VAD classifies the frame as Speech, that signal is output to an SSL module to obtain a better estimate of the desired signal's DOA. (*Id.*)



(*Id.*, Fig. 7 (excerpt).)

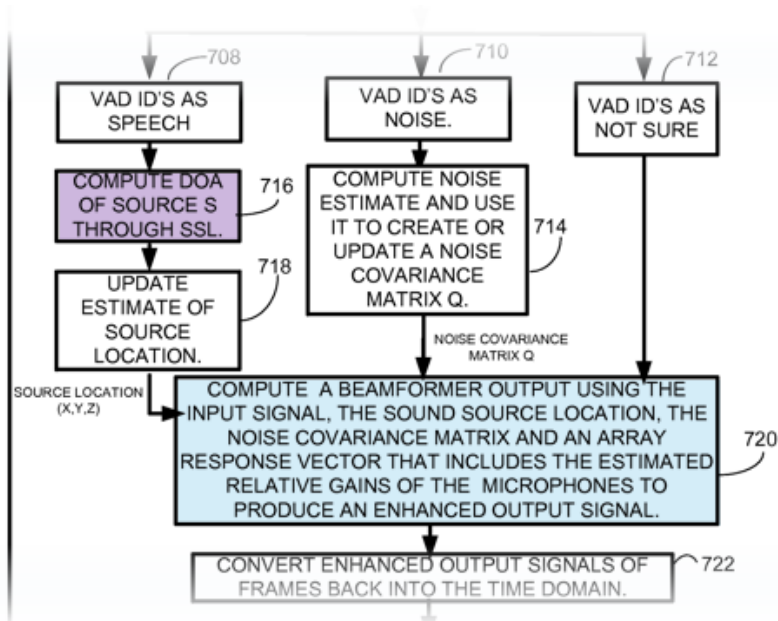
256. Thus, Florencio discloses or renders obvious a speech segment determiner (VAD) configured to [i] determine whether or not a sound picked up by at least either a first microphone or a second microphone (signals from microphones) is a speech segment and [ii] to output (e.g., to SSL module) speech segment information (e.g., whether sound is speech) when it is determined that the sound picked up by the first or the second microphone is the speech segment.

c. Element 1[b]: Voice Direction Detector

257. Element 1[b] recites “a voice direction detector configured, when re-

ceiving the speech segment information, [i] to detect a voice incoming direction indicating from which direction a voice sound travels, based on a first sound pick-up signal obtained based on a sound picked up by the first microphone and a second sound pick-up signal obtained based on a sound picked up by the second microphone and [ii] to output voice incoming-direction information when the voice incoming direction is detected.” Florencio discloses or renders obvious this claim element.

258. Florencio discloses that the VAD output (e.g., whether the sound contains speech) is provided to the SSL module. (EX-1005 (Florencio) ¶¶[0044], [0048], Fig. 7, claim 9.) When the sound contains speech, the SSL module “obtain[s] a better estimate of the location of the desired [e.g., speech] signal which is used in computing the time delay of propagation.” (*Id.* ¶[0044], claim 13.) This is shown in Figure 7:



(*Id.*, Fig. 7 (excerpt).) “The sound source location and received speech frame are

then input into a beamforming module 310[.]” (*Id.* ¶[0044].)

259. Accordingly, Florencio discloses or renders obvious a voice direction detector (e.g., SSL module) configured, when receiving the speech information (e.g., from the VAD), to detect a voice incoming direction indicating from which direction a voice sound travels (e.g., the DOA), based on a first sound pick-up signal obtained based on a sound picked up by the first microphone and a second sound pick-up signal obtained based on a sound picked up by the second microphone (e.g., the signals obtained by the first and second microphones in array) and to output (e.g., to the beamformer) voice incoming-direction information when the voice incoming direction is detected.

d. Element 1[c]: Adaptive Filter

260. Element 1[c] recites “an adaptive filter configured to perform a noise reduction process using the first and second sound pick-up signals based on the speech segment information and the voice incoming-direction information.” Florencio discloses or renders obvious this claim element.

261. Florencio explains that adaptive beamformers (i.e., filters) were known. (EX-1005 (Florencio) ¶¶[0001]-[0004].) Florencio discloses an “enhanced” adaptive beamformer that incorporates a model for directional gains of the sensors and provides a means for estimating those gains. (*Id.* ¶¶[0006], [0025], [0031], [0037]-[0043], [0048].)

262. In addition to the source location and speech frames (e.g., speech signals from the microphone array) (*see* ¶¶257-259, above), Florencio discloses using the sound data that is classified as “Noise” or “Not Sure” in the beamformer. (EX-1005 (Florencio) ¶¶[0044], [0046]-[0048], Figs. 3, 5-7.) The beamformer is adaptive, e.g., parameters like gain are modified based on changing conditions. (*Id.*)

263. Florencio’s adaptive beamformer performs a noise reduction process. (EX-1005 (Florencio) ¶¶[0007] (beamformer output’s an “enhanced output signal with an improved signal to noise ratio”), [0009] (invention provides improved “noise suppression”), [0046], [0048], claim 1; *id.* ¶[0002] (beamforming attenuates noise).) And, it does so using the first and second sound pick-up signals (signals from the first and second microphones in the array). (*Id.* ¶¶[0007] (signals microphone array are used to compute beamformer output), [0045], [0047] (beamformer estimates relative gain of each microphone), [0048] (received signals provided to beamformer after VAD and SSL), Figs. 4-6, claims 1, 6, 9-10, 13.)

264. Florencio’s adaptive beamformer performs noise reduction based on the sound source location information. (EX-1005 (Florencio) ¶¶[0044] (“sound source location” is “input into a beamforming module 310”), [0045] (beamformer uses sound source location to compute array response vector), claims 9 (“computing a beamformer output using ... the sound source location”), 13.)

265. Florencio therefore discloses or renders obvious an adaptive filter (e.g., adaptive beamformer) configured to perform a noise reduction process (increase the signal-to-noise ratio) using the first and second sound pick-up signals based on the speech segment information (e.g., whether signal contains speech) and the voice incoming-direction information (e.g., sound source location).

266. Accordingly, for at least the reasons I discuss above, Florencio discloses or renders obvious each limitation of claim 1, and therefore renders claim 1 obvious as a whole.

2. Claim 13

267. As I discuss above, claim 13 recites a method claim for performing the same limitations as the apparatus of claim 1. (*See* ¶68, above.) Accordingly, Florencio renders obvious claim 13 for the reasons I discuss above for claim 1. (*See* ¶¶253-266, above.)

3. Claims 2 and 14

268. Claim 2 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on a phase difference between the first and second sound pick-up signals.” Claim 14 depends from claim 13 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals.” Florencio discloses or renders obvious these claim elements.

269. Florencio discloses using an SSL algorithm “based on time delay of arrival of the signal and maximum likelihood estimation.” (EX-1005 (Florencio) ¶[0044].) Florencio therefore discloses or renders obvious using the phase difference between the first and second sound pick-up signals to determine the location of the speech signal. (*Id.*)

270. Accordingly, for at least the reasons I discuss above, Florencio discloses or renders obvious each limitation of claims 2 and 14, and therefore renders claims 2 and 14 obvious as a whole.

4. Claims 3 and 15

271. Claim 3 depends from claim 2 and further recites that “the adaptive filter performs the noise reduction process to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first sound pick-up signal has a more advanced phase than the second sound pick-up signal whereas the adaptive filter performs the noise reduction process to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up signal when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal.” Claim 15 depends from claim 14 and further recites that “the noise reduction process is performed to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first

sound pick-up signal has a more advanced phase than the second sound pick-up signal whereas the noise reduction process is performed to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up signal when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal.” Florencio discloses these claim elements.

272. Florencio discloses using a phase difference. (See ¶¶268-270, above.) As I discuss above, a person of ordinary skill in the art would have understood that the first microphone to detect the audio source will be the closest microphone and, therefore, that signal will have a more advanced phase. (See ¶¶199-203, above.) Florencio’s beamformer forms a beam with a high gain in the direction of the audio source, which constructively adds signals from the speaker, and low gain in other directions, which destructively adds signals from the other directions. (See ¶¶260-266, above.)

273. Accordingly, for at least the reasons I discuss above, Florencio discloses or renders obvious each limitation of claims 3 and 15, and therefore renders claims 3 and 15 obvious as a whole.

5. Claim 9

274. Claim 9 depends from claim 1 and further recites that “the speech segment determiner detects the speech segment based on the first sound pick-up signal when the first sound pick-up signal has a more advanced phase than the second

sound pick-up signal whereas the speech segment determiner detects the speech segment based on the second sound pick-up signal when the second sound pick-up signal has a more advanced phase than the first sound pick-up signal.” Florencio renders obvious this claim element.

275. In a microphone array, the microphone closest to the speaker will detect a sound signal with a more advanced phase than microphones at a further distance from the speaker. Florencio analyzes each microphone signal for speech. (EX-1005 (Florencio) ¶[0048], Fig. 7.) Thus, a person of ordinary skill in the art would have understood that Florencio’s VAD detects speech “based on” the first sound pick-up signal (e.g., signal from first microphone) when the first sound pick-up signal has a more advanced phase (e.g., speaker is closer to first microphone) and detects speech “based on” the second sound pick-up signal (e.g., signal from second microphone) when the second sound pick-up signal has a more advanced phase (e.g., speaker is closer to second microphone).

276. Accordingly, for at least the reasons I discuss above, Florencio discloses or renders obvious each limitation of claim 9, and therefore renders claim 9 obvious as a whole.

K. Claims 4, 12, and 16 Would Have Been Obvious in View of Florencio and Visser.

277. Claims 4 and 16 depend from claims 2 and 14, respectively. Furthermore, claim 12 includes limitations that mirror those of claim 1. As I discuss above,

Florencio discloses or renders obvious each limitation of claims 1, 2, and 14. (*See* ¶¶253-270, above.) For at least those reasons and the reasons I discuss below, claims 4, 12, and 16 would have been obvious in further view of Visser. I incorporate my above discussion of Florencio here, and only discuss claim elements for which Visser is relevant in combination with Florencio.

1. Claims 4 and 16

278. Claim 4 depends from claim 2 and further recites that “when the phase difference is within a predetermined range, the adaptive filter outputs either the first or the second sound pick-up signal without performing the noise reduction process.” Claim 16 depends from claim 14 and further recites the same limitation. As I discuss above, Visser discloses or renders obvious these claim elements. (*See* ¶¶124-131, above.)

279. A person of ordinary skill in the art would have been motivated to incorporate Visser’s SSL and noise reduction into Florencio’s device, and would have had a reasonable expectation of success in doing so, for the same reasons I discuss above for incorporating Visser’s SSL and noise reduction into Chen and/or Strömmer. (*See* ¶¶124-131, 208-210, above.)

280. Accordingly, for at least the reasons I discuss above, Florencio and Visser disclose or render obvious each limitation of claims 4 and 16, and therefore render claims 4 and 16 obvious as a whole.

2. Claim 12

281. As I discuss above (*see* ¶¶100-101, above), claim 12 recites two claim elements in addition to those recited in claim 1: element 12[a] recites “a first face and an opposite second face that is apart from the first face with a specific distance” and element 12[b] recites “a first microphone and a second microphone provided on the first face and the second face, respectively.” As I discuss above, Visser discloses these claim elements. (*See* ¶¶146-152, above.)

282. A person of ordinary skill in the art would have been motivated to incorporate Visser’s microphone arrangement into Florencio’s device, and would have had a reasonable expectation of success in doing so, for the same reasons I discuss above for incorporating the microphone arrangement into Chen and/or Strömmer. (*See* ¶¶146-152, 211-213, above.)

283. Accordingly, for at least the reasons I discuss above, Florencio and Visser disclose or render obvious each limitation of claim 12, and therefore render claim 12 obvious as a whole.

L. Claims 5-6 and 17-18 Would Have Been Obvious in View of Florencio and Brandstein.

284. Claims 5-6 and 17-18 depend, directly or indirectly, from claims 1 and 14. As I discuss above, Florencio discloses or renders obvious each limitation of claims 1 and 14. (*See* ¶¶253-270, above.) For at least those reasons and the reasons

I discuss below, claims 5-6 and 17-18 would have been obvious in further view of Brandstein. I incorporate my above discussion of Florencio here.

1. Claims 5 and 17

285. Claim 5 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on magnitudes of the first and second sound pick-up signals.” Claim 17 depends from claim 14 and further recites that “the voice incoming direction is detected based on magnitudes of the first and second sound pick-up signals.” As I discuss above, Brandstein discloses or renders obvious these claim elements. (*See* ¶¶216-227, above.)

286. A person of ordinary skill in the art would have been motivated to incorporate Brandstein’s SRP or SRP-PHAT into Florencio’s device, and would have had a reasonable expectation of success in doing so, for the same reasons as I discuss above for incorporating SRP or SRP-PHAT into Strömmer. (*See* ¶¶216-227, above.)

287. Accordingly, for at least the reasons I discuss above, Florencio and Brandstein disclose or render obvious each limitation of claims 5 and 17, and therefore render claims 5 and 17 obvious as a whole.

2. Claims 6 and 18

288. Claim 6 depends from claim 5 and further recites that “the adaptive filter performs the noise reduction process to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first sound

pick-up signal has a greater magnitude than the second sound pick-up signal whereas the adaptive filter performs the noise reduction process to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up signal when the second sound pick-up signal has a greater magnitude than the first sound pick-up signal.” Claim 18 depends from claim 17 and further recites that “the noise reduction process is performed to reduce a noise component carried by the first sound pick-up signal using the second sound pick-up signal when the first sound pick-up signal has a greater magnitude than the second sound pick-up signal whereas the noise reduction process is performed to reduce a noise component carried by the second sound pick-up signal using the first sound pick-up signal when the second sound pick-up signal has a greater magnitude than the first sound pick-up signal.” Florencio’s beamformer, as modified to include Brandstein’s SRP or SRP-PHAT (as I discuss for claim 5 above, *see* ¶¶285-287, above), discloses or renders obvious these claim elements for the same reasons as the Strömmer-Brandstein combination. (*See* ¶¶228-232, above.)

289. Accordingly, for at least the reasons I discuss above, Florencio and Brandstein disclose or render obvious each limitation of claims 6 and 18, and therefore render claims 6 and 18 obvious as a whole.

M. Claims 7-8 and 19-20 Would Have Been Obvious in View of Florencio, Brandstein, and Visser.

290. Claims 7-8 and 19-20 depend, directly or indirectly, from claims 1, 5, and 14. As I discuss above, Florencio and/or Brandstein disclose or render obvious each limitation of claims 1, 5, and 14. (*See* ¶¶253-270, 285-287, above.) For at least those reasons and the reasons I discuss below, claims 7-8 and 19-20 would have been obvious in further view of Visser. I incorporate my above discussion of Florencio and Brandstein here.

1. Claim 7

291. Claim 7 depends from claim 5 and further recites that “when a power difference that is a difference between magnitudes of the first and second sound pick-up signals is within a predetermined range, the adaptive filter outputs either the first or the second sound pick-up signal without performing the noise reduction process.” As I discuss above, Visser discloses or renders obvious the additional claim element of claim 7. (*See* ¶¶132-135, above.)

292. A person of ordinary skill in the art would have been motivated to incorporate Visser’s SSL and noise reduction into Florencio’s device, and would have had a reasonable expectation of success in doing so, for the same reasons as for incorporating Visser’s SSL and noise reduction into Chen and/or Strömmer. (*See* ¶¶124-131, 234-236, above.)

293. Accordingly, for at least the reasons I discuss above, Florencio, Brandstein, and Visser disclose or render obvious each limitation of claim 7, and therefore render claim 7 obvious as a whole.

2. Claims 8 and 19

294. Claim 8 depends from claim 1 and further recites that “the voice direction detector detects the voice incoming direction based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” Claim 19 depends from claim 14 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” Thus, claims 8 and 19 recite that voice incoming direction is based on a phase difference (as in claims 2 and 14) and magnitude (as in claims 5 and 17). As I discuss above, Visser discloses these claim limitations. (See ¶¶136-145, above.)

295. A person of ordinary skill in the art would have been motivated to use multiple SSL techniques in Florencio’s device, as taught by Visser, and would have had a reasonable expectation of success in doing so, for the same reasons I discuss above for Chen and/or Strömmer. (See ¶¶136-145, 237-240, above.)

296. Accordingly, for at least the reasons I discuss above, Florencio, Brandstein, and Visser disclose or render obvious each limitation of claims 8 and 19, and therefore render claims 8 and 19 obvious as a whole.

3. Claim 20

297. Claim 20 depends from claim 19 and further recites that “the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals and magnitudes of the first and second sound pick-up signals.” I note that the additional claim element of claim 20 is equivalent to the additional claim element of claim 9. (*See* ¶¶96, above.) As I discuss above (*see* ¶¶274-276, above), Florencio renders obvious the additional claim element of claim 20 for the same reasons it renders obvious the additional claim element of claim 9.

298. Accordingly, for at least the reasons I discuss above, Florencio, Brandstein, and Visser disclose or render obvious each limitation of claim 20, and therefore render claim 20 obvious as a whole.

N. Claims 10 and 11 Would Have Been Obvious in View of Florencio, Kim, and Kleffner.

299. Claims 10 and 11 depend from claim 1. As I discuss above, Florencio discloses or renders obvious each limitation of claim 1. (*See* ¶¶253-266, above.) For at least those reasons and the reasons I discuss below, claims 10 and 11 would have been obvious in further view of Kim and Kleffner. I incorporate my above discussion of Florencio here.

300. Claim 10 further recites that “signals are supplied to the voice direction detector as the first and second sound pick-up signals at a sampling frequency of 24 KHz or higher and signals are supplied to the adaptive filter as the first and second sound pick-up signals at a sampling frequency of 12 KHz or lower.” Claim 11 further recites that “the speech segment determiner outputs more accurate speech segment information to the voice direction detector than speech segment information to the adaptive filter.” As I discuss above, Kim and Kleffner disclose these claim elements. (*See* ¶¶156-170, above.)

301. A person of ordinary skill in the art would have been motivated to implement Florencio’s VAD and DOA using a higher (e.g., 48 kHz) sampling frequency as taught by Kim and Kleffner, and then downsample to a lower (e.g., 8 kHz) sampling frequency for Florencio’s beamformer, and would have reasonably expected success, for the same reasons a person of ordinary skill would have been motivated to do so in Chen’s and Strömmer’s systems. (*See* ¶¶156-170, 244-251, above.) For example, it would have improved the accuracy of the VAD and DOA units but reduced the processing requirements for the rest of the process. (*Id.*)

302. Accordingly, for at least the reasons I discuss above, Florencio, Kim, and Kleffner disclose or render obvious each limitation of claims 10 and 11, and therefore render claims 10 and 11 obvious as a whole.

VII. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

303. I am not aware of any secondary considerations of nonobviousness. If the patent owner identifies any alleged evidence of secondary considerations in the future, I reserve the right to respond to that information.

VIII. CONCLUSION

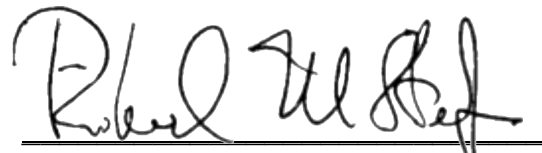
304. For the foregoing reasons, it is my opinion that claims 1-20 of the '259 patent would have been obvious to a person of ordinary skill in the art at the time of the alleged invention in view of the prior art discussed above.

305. I reserve the right to supplement my opinions in the future to address or respond to any issues that the patent owner may raise, as well as new information including, but not limited to, any claim constructions advanced by the patent owner or adopted by the Board in the Institution Decision, and respond to any alleged secondary considerations as they become available to me.

Amazon.com, Inc. v. SoundClear Technologies LLC
Declaration of Richard Stern, Ph.D. – Patent 9,031,259

I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Executed on June 2, 2025 at Pittsburgh, PA.



Richard Stern, Ph.D.