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

PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,031,259

TABLE OF CONTENTS

TABLE OF AUTHORITIES

TABLE OF EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 9,031,259
1002	Declaration of Richard Stern, Ph.D.
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")
1008	Curriculum Vitae of Richard Stern, Ph.D.
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], *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)*, 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., *Adaptive Noise Cancelling: Principles and Applications*, 63 Proc. IEEE 1692 (1975) ("Widrow")
1020	File History of U.S. Patent No. 9,031,259
1021	Declaration of Sylvia Hall-Ellis, Ph.D.

Petitioners Amazon.com, Inc., Amazon.com Services LLC, and Amazon Web Services, Inc. ("Petitioners" or "Amazon") respectfully request *inter partes* review of claims 1-20 of U.S. Patent No. 9,031,259 ("the '259 patent"), which SoundClear Technologies LLC ("Patent Owner" or "PO") purportedly owns.

I. INTRODUCTION

The '259 patent relates to detecting speech and reducing noise in sound signals. The claims require detecting speech, detecting the speech signal's direction, and reducing noise in the signal. These steps were conventional in microphone systems by the patent's earliest possible priority date in 2011 and had been disclosed in many references. Thus, the Board should cancel the claims.

II. BACKGROUND

A. Detecting Speech in Audio Signals Was Known.

Detecting speech in an audio signal is commonly referred to as Voice Activity Detection ("VAD"). (EX-1002 ¶32.) For decades, VAD has been an important part of many areas of speech processing (e.g., automatic speech recognition). (*Id.*) In the 1990s, several standards were adopted for implementing VAD in phones. (*Id.*; EX-1013.) VAD was described in many prior art references. (*E.g.*, EX-1014, 8:30-50 ("VAD functions are well known in the telephony literature"); EX-1010 ¶[0053]; EX-1009 ¶[0101]; EX-1015 ¶[0075]; EX-1005 ¶[0044]; EX-1002 ¶¶32-33.)

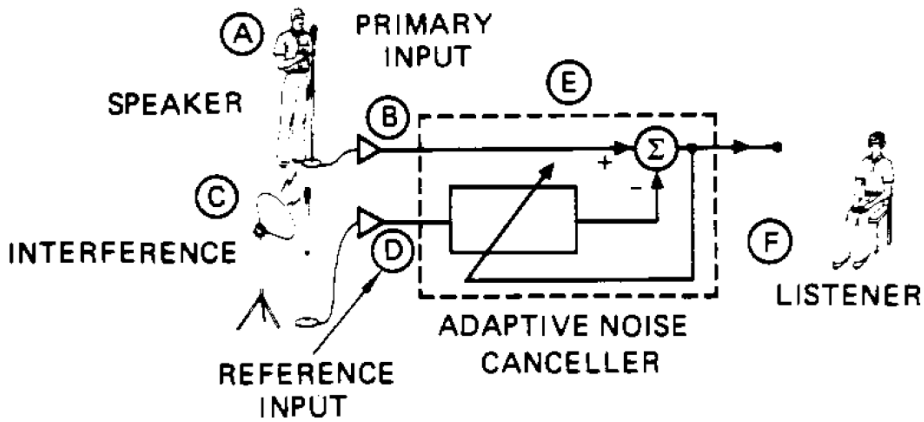
B. Determining an Audio Signal's Incoming Direction Was Known.

Using multiple microphones to determine the direction of an arriving sound signal 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, 157-80; EX-1016, 1043-63.) Many references described determining a sound signal's direction of arrival ("DOA"). (*E.g.*, EX-1010 ¶¶[0053], [0060]-[0063]; EX-1005 ¶[0048]; EX-1014, 7:24-33; EX-1002 ¶34.)

C. Reducing Noise via Adaptive Filters Was Known.

In signal processing, a filter is a device or process that transforms a signal, e.g., by suppressing unwanted components. (EX-1002 ¶35; EX-1017, 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.*)

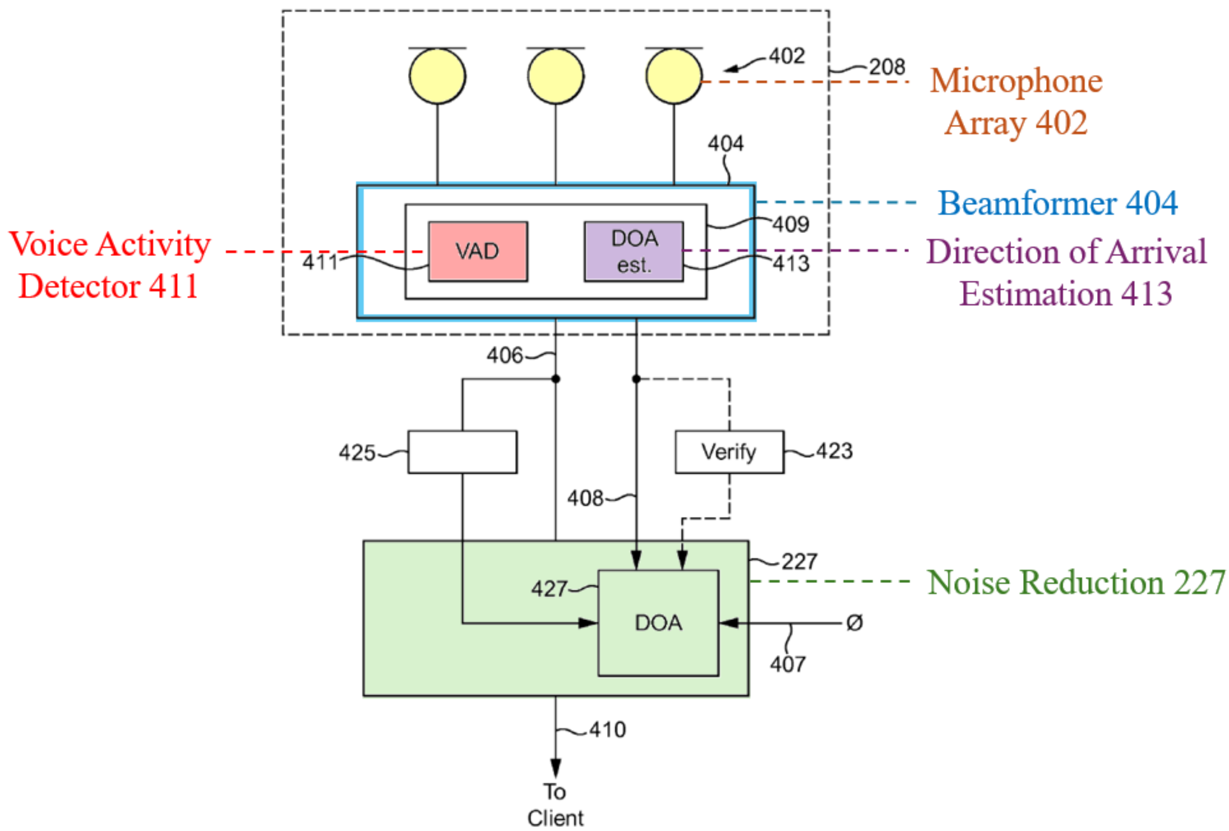
In *adaptive* filters, the parameters are changed over time to optimize the filter's performance. (EX-1018, 14; EX-1002 ¶36.) Using adaptive filters to reduce noise in speech signals has been known since at least 1975, when Stanford University researchers published a seminal article on the topic with the below figure. (EX-1019, 1704.)



(*Id.*) Since then, many references described adaptive filters for noise reduction. (*E.g.*, EX-1011, 266-76; EX-1014, 10:51-11:26; EX-1004, Abstract; EX-1002 ¶36.)

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

Using these three components together 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, Fig. 4[1]; EX-1002 ¶37.)

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

[1] Figures herein may be colored and/or annotated for clarity.

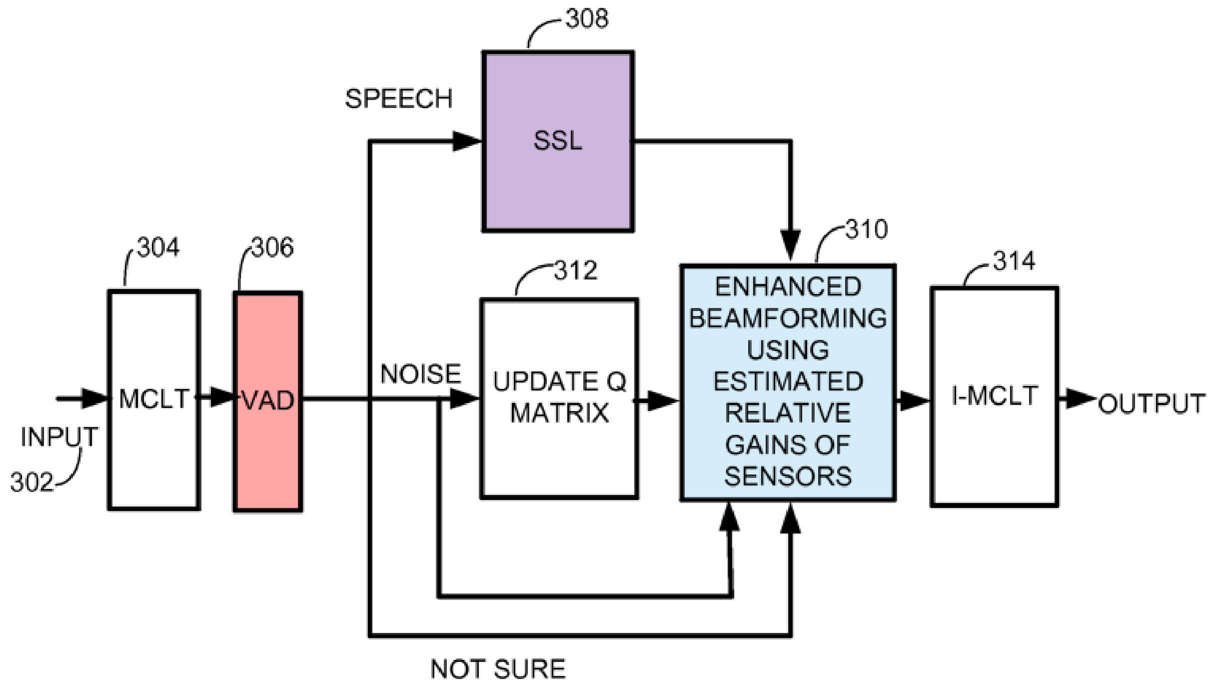


FIG. 3

(EX-1005, Fig. 3; EX-1002 ¶38.)

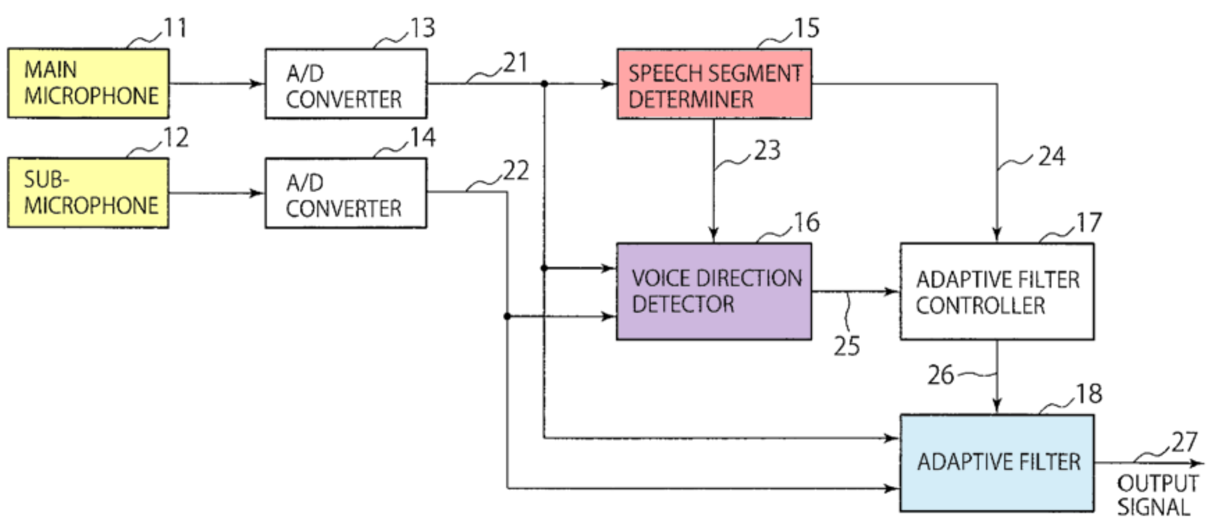
The beamformers in Strömmer and Florencio are adaptive filters because they automatically optimize their filtering behavior over time. (EX-1002 ¶39.) A beamformer is a spatial filter: it emphasizes signal components from a desired direction while suppressing components from other directions. (EX-1005 ¶[0008]; EX-1016, 946; EX-1002 ¶39.) By focusing a beamformer in a talker's direction, that person's voice can be emphasized and noise from other directions can be filtered out. (*Id.*)

As Strömmer acknowledges, adaptive beamformers were already well known. (EX-1010 ¶¶[0008]-[0009].) Adaptive beamformers can change the focus direction based on a speech signal's estimated DOA. (*Id.*) Thus, the adaptive beamformers in Strömmer and Florencio are adaptive filters that perform noise reduction. (EX-1002 ¶40; *see* EX-1011, 89 (beamformer as "adaptive noise canceler").)

III. THE '259 PATENT

A. Overview

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



(*Id.*, Fig. 1; EX-1002 ¶41.)

The patent admits that the three signal processing components were known. (EX-1001, 4:48-49 ("speech segment determiner" can employ "any speech segment determination techniques"), 16:66 ("several techniques for voice direction detection" exist); 23:59-24:8 (adaptive filter performs "a regular noise reduction process"); EX-1002 ¶42.)

B. Prosecution History

Despite an abundance of prior art, the Examiner allowed the claims without a single Office Action. (EX-1020, 30-37.) None of the references relied on herein were submitted to or considered by the Examiner.

C. Priority

The '259 patent claims priority to foreign applications filed on September 15, 2011. (EX-1001, 1.) Petitioners do not concede that the claims are entitled to the foreign priority date.

IV. STATEMENT OF PRECISE RELIEF REQUESTED

A. Grounds

Petitioners request cancellation of claims 1-20 under 35 U.S.C. §103 as follows:

Ground	Reference(s)	Challenged Claims
1A	Chen and Li	1-3, 5-6, 8-9, 12-15, 17-20
1B	Chen, Li, and Florencio	1-3, 5-6, 8-9, 12-15, 17-20

Ground	Reference(s)	Challenged Claims
1C	Grounds 1A or 1B and Visser	4, 7-8, 12, 16, 19-20
1D	Grounds 1A or 1B, Kim, and Kleffner	10-11
2A	Strömmer	1-3, 9, 13-15
2B	Strömmer and Visser	4, 12, 16
2C	Strömmer and Brandstein	5-6, 17-18
2D	Strömmer, Brandstein, and Visser	7-8, 19-20
2E	Strömmer, Kim, and Kleffner	10-11
3A	Florencio	1-3, 9, 13-15
3B	Florencio and Visser	4, 12, 16
3C	Florencio and Brandstein	5-6, 17-18
3D	Florencio, Brandstein, and Visser	7-8, 19-20
3E	Florencio, Kim, and Kleffner	10-11

Additional support is included in the Declaration of Dr. Richard Stern. (EX-1002.)

B. Status of References as Prior Art

The following references are prior art under pre-AIA §102(b) because each published more than one year before the ’259 patent’s U.S. filing date. M.P.E.P. §2151.

Reference	Publication Year	Exhibit
Chen	2010	EX-1003
Florencio	2008	EX-1005

Reference	Publication Year	Exhibit
Kim	2010	EX-1006
Kleffner	2009	EX-1007
Brandstein	2001	EX-1011

The non-patent references were publicly available more than one year before the '259 patent's U.S. filing date. (EX-1021.)

The following references are prior art under pre-AIA §102(e):

Reference	Filing Date	Exhibit
Li	December 15, 2010	EX-1004
Strömmer	August 18, 2011	EX-1010
Visser	February 17, 2011	EX-1012

The references are analogous art because each is from the same field of endeavor as the '259 patent, e.g., signal processing and/or automatic speech recognition. (EX-1002 ¶22.) They are also pertinent to a particular problem the inventor was focused on, e.g., reducing noise in audio signals containing speech. (*Id.*)

V. LEVEL OF ORDINARY SKILL

A POSITA would have had at least 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. (EX-1002 ¶¶27-31.) Work experience could substitute for formal education and additional formal education could substitute for work experience. (*Id.*)

VI. CLAIM CONSTRUCTION

No claim terms require construction to resolve the invalidity challenges here. *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co. Ltd.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017); *Vivid Techs., Inc. v. Am. Sci. & Eng'g, Inc.,* 200 F.3d 795, 803 (Fed. Cir. 1999). For purposes of this proceeding only, Petitioners assume the claims are not invalid under §112.

VII. GROUND 1A: CLAIMS 1-3, 5-6, 8-9, 12-15, AND 17-20 WOULD HAVE BEEN OBVIOUS IN VIEW OF CHEN AND LI.

Chen discloses a device (e.g., mobile phone) for enhancing voice quality in a signal. (EX-1003 ¶[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].)

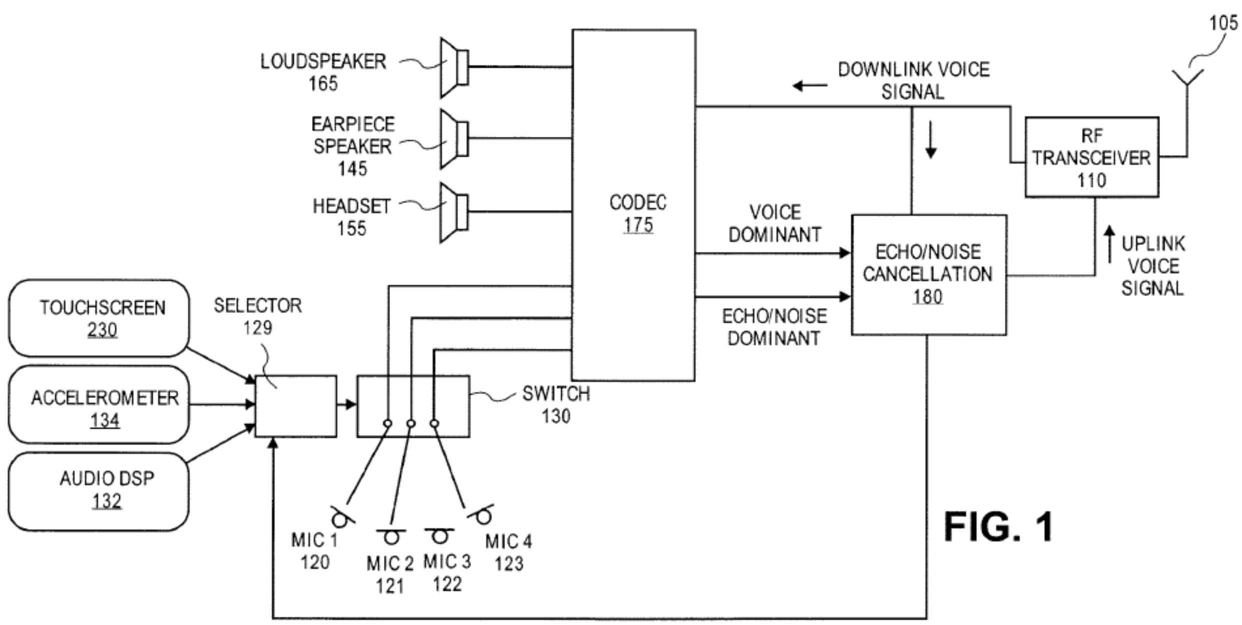


FIG. 1

(*Id.*, Fig. 1.) Selector 129 selects one microphone as the "primary" microphone to provide a *voice*-dominant signal to codec 175, and selects another microphone 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 voice signal's DOA (*id.* ¶[0019]). (EX-1002 ¶44.) The voice-dominant and noise-dominant signals are input to an echo/noise-cancellation controller 180, which "refines the voice dominant signal … by suppressing noise and cancelling echo with the assistance of one or more signals from the other selected microphones." (*Id.* ¶¶[0022], [0028], Fig. 3.)

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, as shown:

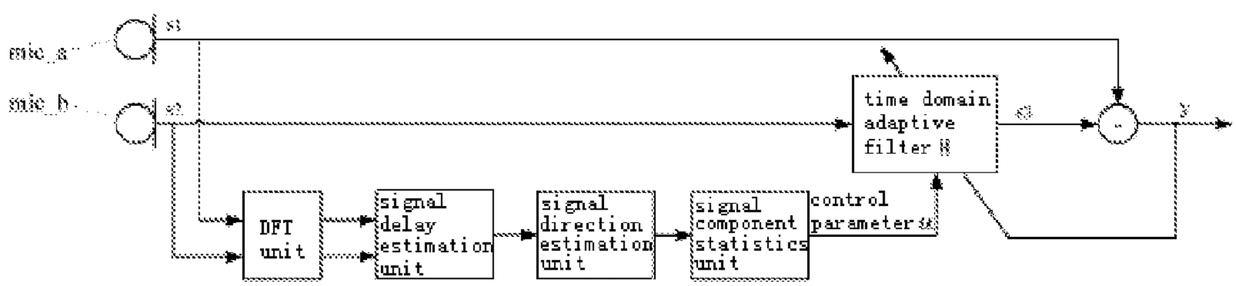


Figure 4

(EX-1004, Fig. 4, Abstract, ¶¶[0010], [0048]; EX-1002 ¶45.)

A. Claim 1

1. Preamble

Chen discloses a "noise reduction apparatus" because it discloses a handheld device that enhances a "voice dominant signal" by "suppressing ambient noise or canceling echo therein." (EX-1003, Abstract, Figs. 1-4; EX-1002 ¶46.) An exemplary apparatus is the phone shown in Chen's Figure 2:

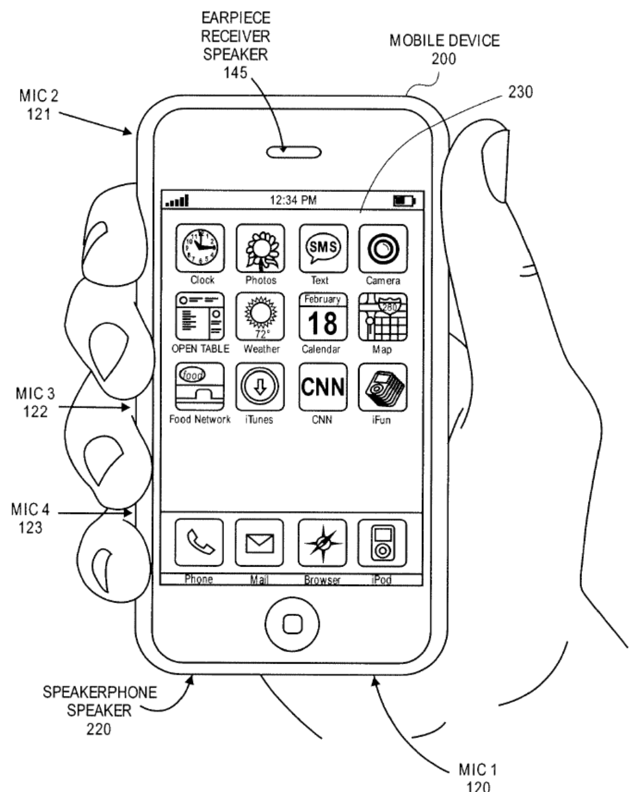


FIG. 2

(EX-1003, Fig. 2.)

2. Element 1[a]

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 limitation. (EX-1002 ¶¶47-51.) Chen discloses a DSP 132 that "monitors (to compare or analyze) the available microphone signals and provides the results of its analysis to a selector." (EX-1003 ¶[0015].) Chen's selector selects the voice-dominant signal based on, e.g., the "signal-to-noise ratio" (commonly referred to as "SNR"). (*Id.* ¶[0016]; EX-1002 ¶48.) A POSITA would have understood this disclosure to be a reference to determining whether each microphone signal contains a speech segment because Chen describes selecting a "voice dominant signal" in a mobile phone via SNR, and SNR was a standard way of performing VAD in mobile phones at the time. (*Id.*; EX-1013, 18 (defining SNR), 22 (SNR used in VAD decision).)

Even if Chen did not disclose this limitation, it would have been obvious over Chen. The purpose of Chen's selector 129 is to select the best signal as the voice-dominated signal. A POSITA would have immediately recognized that using VAD to identify which 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 contains voice. (EX-1002 ¶49.) A VAD also would have helped the selector select, as the echo/noise-dominant signal, a signal without speech where possible. (*Id.*)

Chen discloses that the DSP output is provided to the selector. (EX-1003 ¶¶[0015], [0034].)

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. (EX-1002 ¶¶47-51.)

3. Element 1[b]

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 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 ¶[0019].) A POSITA would have understood that this angle identification (i.e., DOA determination) would be performed by DSP 129 or processor 704, and Chen discloses providing the result to the selector to choose the best signal. (EX-1002 ¶53.)

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 microphones 120, 121) and to output voice incoming-direction information when the voice incoming direction is detected (e.g., to the selector). (*Id.* ¶¶52-54.)

4. Element 1[c]

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 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 ¶[0022]; *id.* ¶¶[0028], [0034].) The echo/noise cancellation controller 180 outputs "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; EX-1002 ¶56.) Chen does not expressly refer to its noise cancellation module as an "adaptive filter."

Li describes an improved adaptive filter for noise reduction in "[w]ireless mobile communication … devices." (EX-1004 ¶[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):

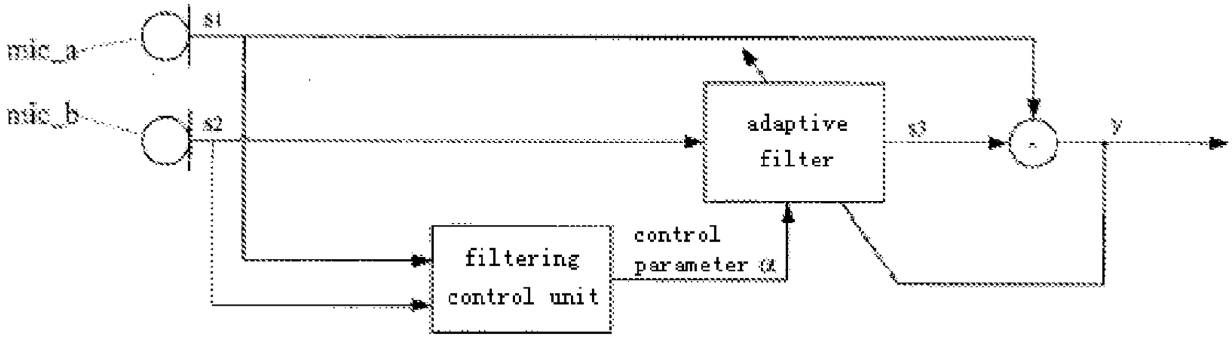


Figure 2

(*Id.*, Fig. 2, ¶¶[0010], [0036], [0045]-[0048]; *id.*, claim 1 ("noise reduction method … comprising … adaptive filter").) Li confirms the widely known use of adaptive filters for noise reduction. (*Id.* ¶[0045] ("typically, noise reduction is implemented using an adaptive filter"); EX-1002 ¶57.)

When combined with Chen, Li's adaptive filter would perform noise reduction 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. (EX-1002 ¶58.) First, as discussed, Chen uses speech segment information 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. (*Id.*) 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 ¶¶[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); EX-1002 ¶58.) Thus, the adaptive filtering would be "based on" the speech segment information.

Li's adaptive filter also performs noise reduction based on the voice incoming-direction information. (EX-1002 ¶59.) First, as previously described, Chen uses the voice incoming direction information to select the voice-dominant signal ("desired voice signal" in Li) to be enhanced by Li's adaptive filter. (*Id.*) Second, Li's adaptive filter uses the "incidence angle" of the incoming voice signals to control the adaptive filter. (EX-1004 ¶[0069], claims 1-6, 10, 11; EX-1002 ¶59.)

Accordingly, Li discloses or renders obvious the adaptive filter recited in element 1[c]. (EX-1002 ¶¶55-60.)

5. Motivation to Combine

A POSITA would have been motivated to incorporate Li's improved adaptive filter into Chen's echo/noise cancellation stage, for many reasons. (*Id.* ¶¶61-66.)

First, Chen performs noise reduction to enhance a voice-dominant signal using a noise/echo-dominant signal (*e.g.*, EX-1003 ¶[0015], claim 7), but does not provide details of that noise reduction stage. A POSITA would have looked to other references, such as Li, to understand how to implement the noise reduction step. (EX-1002 ¶62.) Li teaches doing so using an adaptive filter. (*Id.*)

Second, Li's adaptive filter eliminates noise, enhances SNR, and protects speech quality. (EX-1004 ¶[0010].) A POSITA would have been motivated to modify Chen's system to include Li's adaptive filter to obtain these benefits. (EX-1002 ¶63.)

Third, reducing noise via adaptive filters was widely known. (EX-1004 ¶[0045]; *supra* §II.C; EX-1002 ¶64). Thus, implementing Li's adaptive filter in Chen's system would involve the simple substitution of one known element (Li's adaptive filter) for another known element (Chen's noise reducer) to obtain predictable results (improved noise reduction). (EX-1002 ¶64); *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 417 (2007). It would also reflect using a known technique (Li's adaptive filter) to improve a similar device and method (Chen's) in the same way (to enhance signal quality). (*Id.*) 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). (*Id.*)

A POSITA would have reasonably expected success in modifying Chen in this way. Doing so would have been trivial, involving configuring Chen's noise reduction stage with an adaptive filter as was widely known. (EX-1002 ¶65.) Further, Li implements its adaptive filter in a very similar system. (*Id.*; EX-1003, Abstract; EX-1004 ¶[0002].)

Accordingly, claim 1 would have been obvious over Chen and Li. (EX-1002 ¶¶46-66.)

B. Claim 13

Independent claim 13 is a method claim with limitations corresponding to the limitations in claim 1. It is therefore unpatentable for the same reasons as claim 1. (*Id.* ¶¶67-69.)

C. Claims 2 and 14

Dependent claim 2 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 adds the same limitation to claim 13. The "phase difference" refers to the "amount of delay" between the audio signals detected by the first and second microphones. (EX-1001, 18:41-47.)

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 delays in a "signal direction estimation unit for calculating incidence angles of the microphone array signals[.]" (EX-1004 ¶[0030]; *id.* ¶[0053] (determine target speech signal angles based on "phase differences"); EX-1002 ¶72.) 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. (EX-1002 ¶72.)

A POSITA 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. (*Id.* ¶¶73-79.)

First, Chen discloses determining the DOA "by measuring for the maximum audio signal[.]" (EX-1003 ¶[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 ¶¶[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 POSITA would have understood that implementing Li's adaptive noise reduction filter as described above for claim 1 would include Li's voice-angle determination methodology, which is based on phase differences. (EX-1002 ¶74.) A POSITA 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." (*Id.*)

Second, while Chen discloses determining the DOA 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 ¶[0016].) A POSITA would have understood that determining the DOA based on both power (as taught by Chen) and phase difference (as taught by Li) would lead to a more accurate DOA determination and thus improved selection of the "best" microphone signal. (EX-1002 ¶75.)

Third, a POSITA 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-1002 ¶76; EX-1012, 28:53-29:34; *Infra* §IX.C.)

Fourth, determining the DOA using phase differences in Chen's system would merely represent the addition of a known element (phase-based DOA determination) to another known element (Chen's DOA determination) to obtain predictable results (more accurate DOA determination). (EX-1002 ¶77); *KSR*, 550 U.S. at 417. The combination would also reflect using a known technique (phase-based DOA determination) to improve a similar device and method (Chen's) in the same way (to identify and enhance desired signals). (*Id.*) It would also reflect applying a known technique (phase-based DOA determination) to a known device and method (Chen's) that is ready for improvement and yields predictable results (improved DOA determination for improved signal selection). (*Id.*)

A POSITA would have reasonably expected success in using phase-based DOA (as Li describes) in Chen's device because phase-based DOA techniques were widely known, routinely implemented, and Li teaches implementing such techniques on a similar device. (EX-1002 ¶78.)

Thus, Li discloses or renders obvious the additional limitation of claims 2 and 14, and the claims as a whole would have been obvious over Chen and Li. (*Id.* ¶¶70-79.)

D. Claims 3 and 15

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, which depends from claim 14, recites the same limitation. Chen and Li render this limitation obvious.

Chen discloses selecting, from multiple microphone signals, one as the voice-dominant signal and one as the noise/echo-dominant signal. (EX-1003 ¶[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 switching between microphones to always select the best microphone while the target speaker moves. (*Id.* ¶¶[0019], [0025]; EX-1002 ¶81.) Li teaches identifying the target speaker's direction using phase differences. (*Supra* §VII.C.)

A POSITA would have understood that the Chen-Li combination using phase-based DOA satisfies this claim limitation. With reference to Chen's Figure 2, 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. (EX-1002 ¶82.)

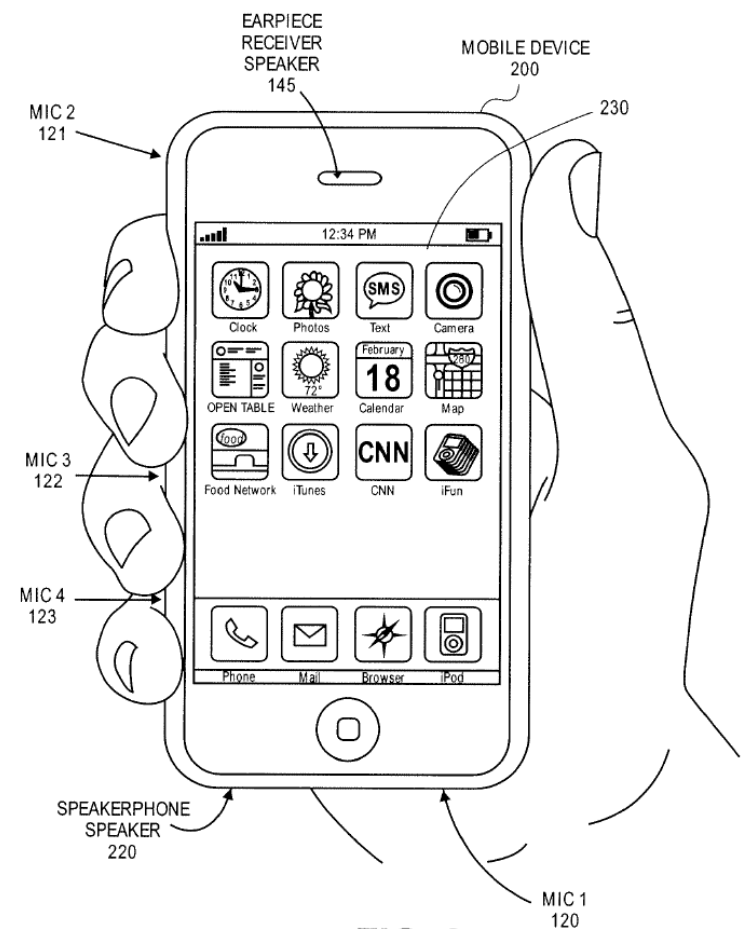


FIG. 2

(EX-1003, Fig. 2.)

If the target speaker then moves around the device (or changes the device orientation) such that the incoming voice direction is 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. (*Id.* ¶¶[0019], [0004] (switching of primary microphone may "occur dynamically" when talker "moves from one side of the device to another"); EX-1002 ¶83.)

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., 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., talker is closer to MIC 2 than MIC 1). (EX-1002 ¶¶80-85.)

E. Claims 5 and 17

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 recites a similar limitation.

Chen discloses detecting the voice incoming direction (angle to target speaker) "by measuring for the maximum audio signal picked up for the target speaker." (EX-1003 ¶[0019].) A POSITA would have understood Chen's reference to "maximum audio signal" to refer to the magnitude (or "power") of the signals. (EX-1002 ¶87.) Moreover, Chen discloses choosing the best microphone signal based on, *inter alia*, which signal is the "highest powered or loudest signal," and which has the highest "signal-to-noise ratio." (EX-1003 ¶[0016].) Thus, Chen discloses or renders obvious this limitation. (EX-1002 ¶¶86-88.)

F. Claims 6 and 18

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, which depends from claim 17, recites the same thing.

As discussed for claim 5, Chen discloses or renders obvious determining the DOA 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 ¶¶[0015]-[0016].) And, the device dynamically switches which microphone signal is the voice-dominant signal. (*Id.* ¶¶[0004], [0019], [0026]; EX-1002 ¶90.)

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 and is therefore selected as voice-dominant signal). (EX-1002 ¶¶89-92.)

G. Claims 8 and 19

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). Li discloses determining the DOA based on a phase difference. (*Supra* §VII.C.) Chen discloses determining the DOA based on the magnitudes. (*Supra* §VII.E.) A POSITA would have been motivated to incorporate Li's phase-based DOA determination into Chen's device and would have reasonably expected success in doing so as discussed for claim 2. (*Supra* §VII.C.) For example, using both DOA techniques would improve Chen's DOA determination, facilitate the use of Li's adaptive filter, be consistent with Chen's teaching that microphone selection should not be based on any one factor, but on a combination of various factors, and allow the device to select or weigh the appropriate DOA technique for the circumstances. (EX-1002 ¶94; EX-1012, 28:53-29:34; *see infra* §IX.C.)

Thus, claims 8 and 19 would have been obvious over Chen and Li. (EX-1002 ¶¶93-95.)

H. Claims 9 and 20

Claim 9 depends from claim 1 and recites detecting 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, and vice versa. In other words, detecting speech based on the sound signal from the microphone nearest to the person talking. (EX-1002 ¶96.) Claim 20, which depends from claim 19, recites the same limitation.

In a two-microphone system, Chen analyzes both microphone signals for speech (e.g., signal-to-noise ratio). (EX-1003 ¶¶[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., 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., talker is closer to MIC 2). (EX-1002 ¶97.)

Thus, Chen and Li disclose, or at least renders obvious, claims 9 and 20. (*Id.* ¶¶96-98.)

I. Claim 12

Claim 12 recites limitations similar to claim 1, but adds "a first face and an opposite second face that is apart from the first face with a specific distance" and "a first microphone and a second microphone provided on the first face and the second face, respectively."

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

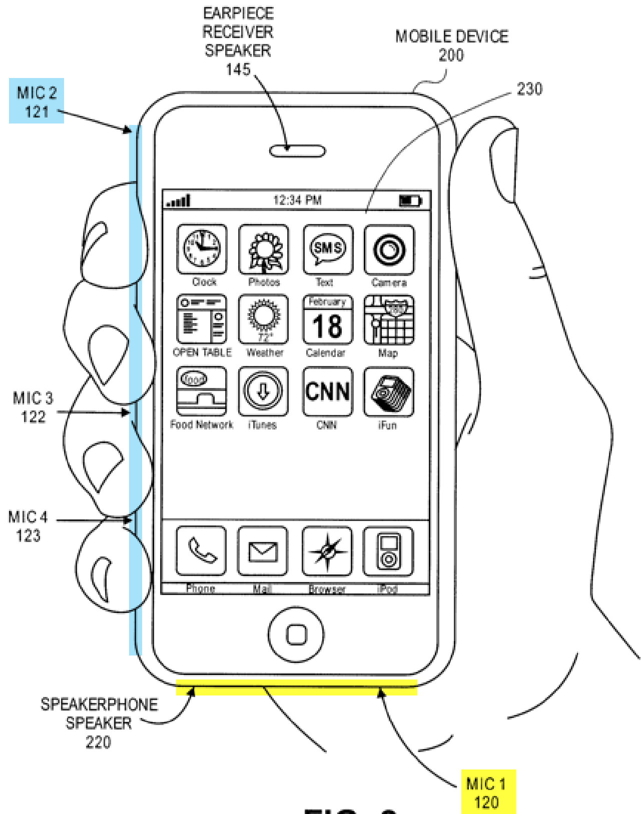


FIG. 2

(EX-1003, Fig. 2.) However, Chen teaches that the microphones may be "located anywhere on the device." (*Id.* ¶[0024]; *id.* ¶[0028].) Thus, Chen discloses or renders obvious microphones on opposing faces (e.g., front and back, top and bottom, left and right sides). (EX-1002 ¶102.) Moreover, Chen discloses that one microphone may face an interviewee while another microphone faces the interviewer. (EX-1003 ¶[0018].) This would, at a minimum, suggest placing one microphone on the back of the device and another on the front to improve signal quality during interviews. (EX-1002 ¶102.)

Thus, Chen discloses or renders obvious the additional limitation of claim 12, and claim 12 would have been obvious in view of Chen and Li. (*Id.* ¶¶99-103.)

VIII. GROUND 1B: CLAIMS 1-3, 5-6, 8-9, 12-15 AND 17-20 WOULD HAVE BEEN OBVIOUS IN VIEW OF CHEN, LI, AND FLORENCIO.

These claims also would have been obvious over Chen and Li (as in Ground 1A) in view of Florencio. (EX-1002 ¶104.) Petitioners incorporate the discussion of Chen and Li from Ground 1A, and therefore address here only the limitations to which Florencio pertains.

Florencio describes a noise reduction system for devices such as mobile phones. (EX-1005 ¶[0019].) The system includes a microphone array and a VAD (306) to determine whether the signals contain speech:

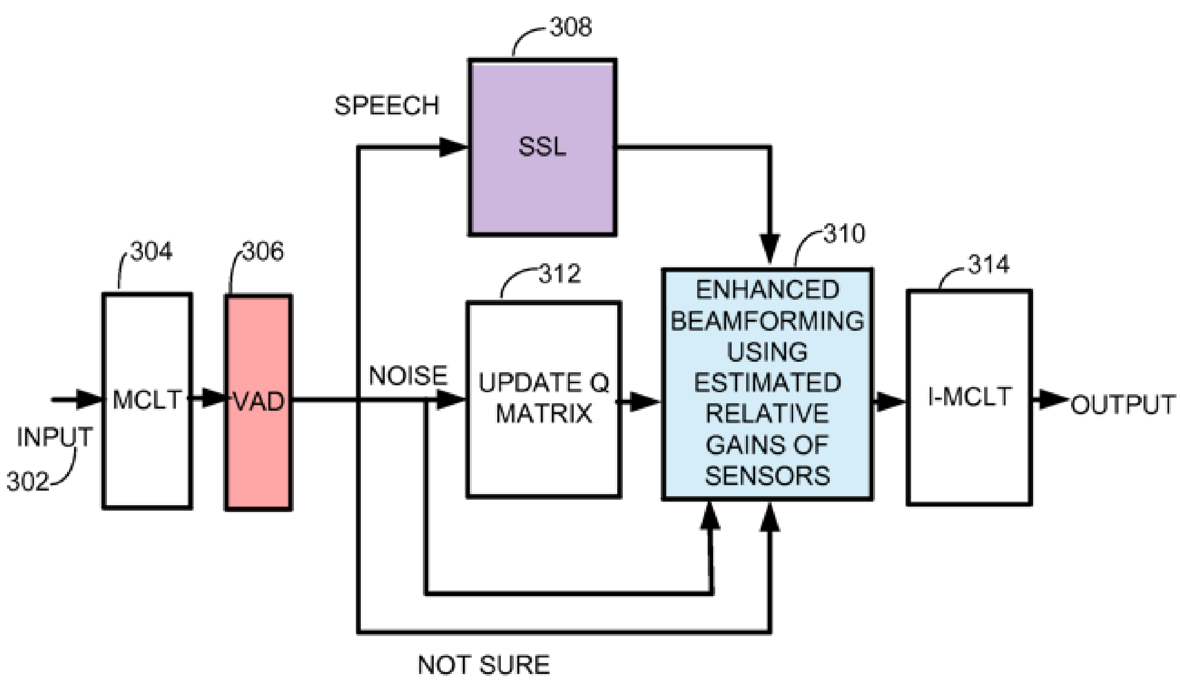


FIG. 3

(*Id.*, Fig. 3, ¶¶[0020]-[0024], [0044]-[0045], [0048], Figs. 1-7.) Sound classified as speech is processed (e.g., by SSL 308) to determine the source location. (*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:

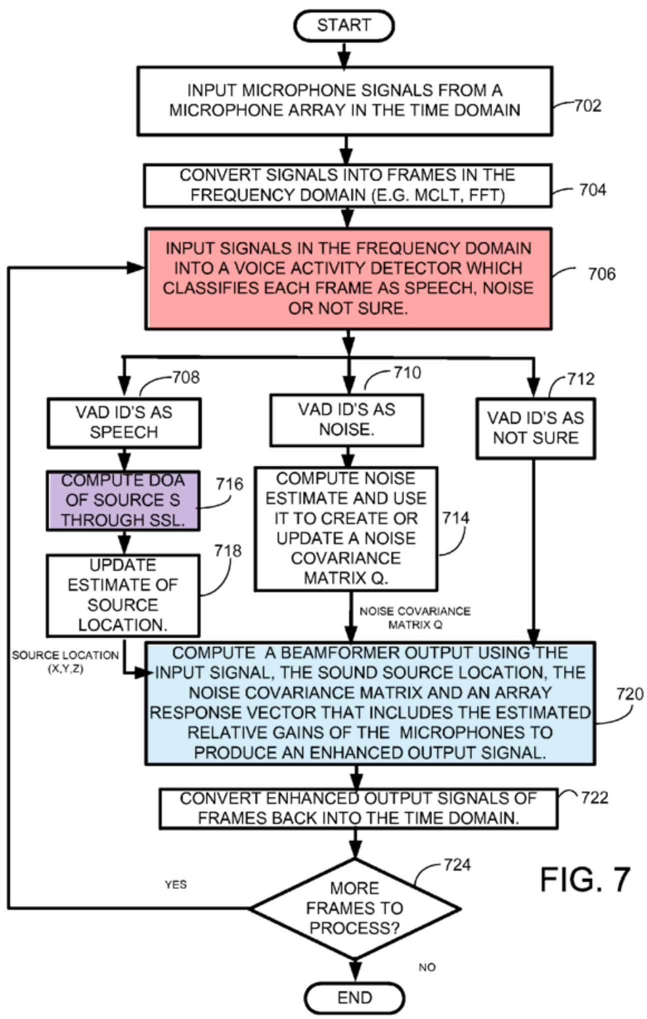


FIG. 7

(*Id.*, Fig. 7; EX-1002 ¶105.)

A. Claims 1 and 13

As shown and as discussed in more detail in Ground 3, 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. (EX-1002 ¶106; *infra* §XVI.A.2.) 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.*; *infra* §§XVI.A.2-XVI.A.3.) And, as discussed in Ground 1A, Li's adaptive filter performs noise reduction based on the speech segment information and the voice incoming-direction information. (EX-1002 ¶106.)

To the extent PO argues 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, these limitations would have been obvious in view of Florencio. A POSITA would have been motivated to modify the Chen-Li combination to include a VAD and SSL unit, as taught by Florencio. (*Id.* ¶¶107-13.)

First, Chen teaches that the microphone signals are analyzed and compared by, e.g., a DSP. (EX-1003 ¶¶[0015], [0034].) The selector (129 or 732) then identifies the "best" available signal 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 "various factors" may be considered. (*Id.* ¶[0016].) A POSITA would have recognized that analyzing the signals to determine whether they 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 ¶[0048].) A POSITA would have immediately recognized that such a VAD would help Chen's selector by ensuring that the signal selected as the voice-dominant signal contains speech. (EX-1002 ¶108.)

Second, Chen discloses using "audio tracking or audio beam forming" to identify the DOA. (EX-1003 ¶[0019].) Consequently, a POSITA would have been motivated to look to other references, like Florencio, that describe beamforming methods involving DOA determination and its implementation. (EX-1002 ¶109.) And, especially in the scenario Chen describes regarding a moving target, both VAD and SSL would help the selector identify the best signal to use at any given time. (*Id.*)

Third, the use of VADs and SSL algorithms in beamformers was widely known, as the patent admits and many prior art references demonstrate. (*Supra* §II.) Thus, implementing VAD and SSL (as taught by Florencio) in Chen's microphone signal analyzer would merely represent the simple addition of known elements

(VAD and SSL 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). (EX-1002 ¶110); *KSR*, 550 U.S. at 417. The combination would reflect using known techniques (VAD and SSL) to improve a similar device (Chen's) in the same way (determining signals with speech and their direction). (*Id.*) It would also apply known techniques (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). (*Id.*)

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, claim 1, ¶¶[0065], [0069], [0020].) Thus, a POSITA 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. (EX-1002 ¶111.)

A POSITA 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 a VAD and SSL unit, as taught by Florencio. (*Id.* ¶112.)

Accordingly, claims 1 and 13 would have been obvious over Chen, Li and Florencio. (*Id.* ¶¶106-14.)

B. Claims 2 and 14

Florencio discloses using an SSL algorithm "based on time delay of arrival of the signal and maximum likelihood estimation." (EX-1005 ¶[0044].) Florencio therefore also discloses or renders obvious using the phase differences to determine the DOA as recited in claims 2 and 14. (*Infra* §XVI.C; EX-1002 ¶¶115-17.)

C. Claims 5 and 17

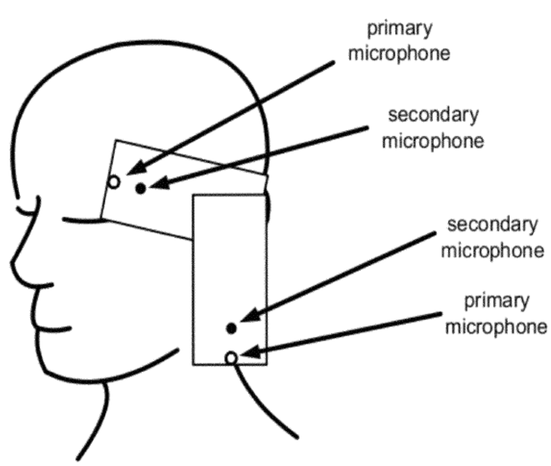
Chen and Li render these claims obvious. (*Supra* §VII.E.) They also would have been obvious in view of Florencio. In view of Chen's disclosure, a POSITA implementing Florencio's SSL unit would have been motivated to determine the DOA based at least in part on the magnitude of the signals. (EX-1002 ¶119.) 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 ¶[0044]), which a POSITA would have understood to refer to the "magnitude" of the signals in each direction. (EX-1002 ¶119.) 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." (*Id.* ¶¶118-20.)

IX. GROUND 1C: CLAIMS 4, 7-8, 12, 16, AND 19-20 WOULD HAVE BEEN OBVIOUS IN VIEW OF THE REFERENCES IN GROUND 1A OR 1B AND VISSER.

A. Claims 4 and 16

Claim 4, which depends from claim 2, recites "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 recites the same limitation. These claims would have been obvious over the references in Grounds 1A and 1B and Visser. (EX-1002 ¶¶124-31.)

Visser discloses a system to reduce noise in audio signals detected by a microphone array. (EX-1012, 7:22-31.) Visser explains that, when an audio source is "broadside" to the array as in Figure 2 below (i.e., far from an axis of the microphone array), "dual-microphone noise reduction may not be possible" because the detected sound signals "are basically very similar" and, as a result, dual-microphone noise reduction may attenuate the desired voice signal. (*Id.*, 7:8-17, 7:49-54.)



(*Id.*, Fig. 2.)

In this scenario, Visser does not perform “dual-microphone noise reduction,” but rather “single-microphone noise reduction.” (*Id.*, 7:22-25; EX-1002 ¶¶125-26.) 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.” (EX-1012, 5:15-20.) A POSITA 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. (EX-1002 ¶126.)

Thus, Visser discloses, or at least renders obvious, claims 4 and 16. (*Id.* ¶¶124-26.)

A POSITA would have been motivated to modify the Chen-Li combination to include Visser’s single-signal processing option for many reasons. (*Id.* ¶¶127-31.)

First, Visser explains that when the audio signal is broadside to the array, dual-microphone noise reduction will attenuate the desired speech signal. (EX-1012, 7:14-17.) Chen and Li disclose dual-microphone systems. Thus, a POSITA would have been motivated to implement one solution Visser describes, e.g., to not use dual-microphone noise reduction and to instead output one of the signals. (EX-1002 ¶128.) A POSITA 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. (*Id.*)

Second, the use of SSL techniques and dual- and single-microphone noise reduction were widely known, as Visser and other references demonstrate. (EX-1012, 1:51-55; *supra* §II.) Thus, implementing single-signal noise reduction (as taught by Visser) in the combined Chen-Li 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 signal when the DOA is within a certain broadside range). (EX-1002 ¶129); *KSR*, 550 U.S. at 417. 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). (*Id.*) 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). (*Id.*)

A POSITA 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. (EX-1002 ¶130.)

B. Claim 7

Claim 7, which depends from claim 5, recites "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." Claim 7 is similar to claim 4 but it uses a "difference of magnitude" rather than a "difference of phase."

As described for claim 5, Chen discloses using magnitude to determine DOA. (*Supra* §VII.E.) So does Visser. (EX-1012, 6:57-61 (DOA methods include "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. (*Supra* §IX.A.)

A POSITA 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. (*Supra* §IX.A.) Thus, a POSITA would have known to set a predetermined range of magnitude differences that would indicate the speaker is broadside. (EX-1002 ¶134.) In other words, if the difference in signal magnitudes was within a certain range, the speaker is broadside to the microphones. (*Id.*) In that scenario, Visser discloses outputting either microphone signal without performing the dual-microphone noise reduction. (*Supra* §IX.A.) Thus, Visser discloses or renders obvious claim 7, and the claim as a whole would have been obvious over the references in Grounds 1A or 1B and Visser. (EX-1002 ¶¶132-35.)

C. Claims 8 and 19

Claims 8 and 19 recite that voice incoming direction is determined based on a phase difference (as in claims 2 and 14) *and* magnitude (as in claims 5 and 17). The references in Grounds 1A and 1B render obvious claims 8 and 19. (*Supra* §§VII-VIII.) Those claims also would have been obvious in view of Visser.

Visser discloses using both phase differences and magnitude differences for DOA determination. Specifically, Visser discloses determining "directions of arrival (DOAs)" "based on differences in phase and/or level (e.g., amplitude, gain, energy) between the channels." (EX-1012, 6:57-61; *id.*, 28:53-29:3 (DOA information may be "based on phase differences" and "additionally … based on gain differences"), 29:8-13 (system uses "phase-difference-based processing … at some times, and us[es] gain-difference-based processing at other times"), 29:20-34 ("directional indications from phase-difference-based and gain-difference-based processing techniques" may be "combin[ed]"); EX-1002 ¶137.) Thus, Visser discloses this limitation. (EX-1002 ¶¶136-37.)

A POSITA 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. (*Id.* ¶¶138-45.)

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, 29:4-13.) Thus, a POSITA would have understood it would be beneficial to switch between phase-based and magnitude-based DOA techniques "depending on the … estimated distance between source and microphone." (*Id.*; EX-1002 ¶139.)

Second, Visser describes selecting microphone signals by combining DOA indications from both phase-based and gain-based processing techniques, e.g., 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, 29:20-34.) A POSITA 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. (EX-1002 ¶140.)

Third, Chen and Visser both describe methods for determining DOA and reducing noise in mobile phones. (EX-1003 ¶[0012]; EX-1012, Fig. 1, 6:41-49; EX-1002 ¶141.)

Fourth, in view of the discussion in Chen (and Li and Florencio) regarding determining the DOA, a POSITA would have been motivated to look at other references for improved DOA techniques. (EX-1002 ¶142.) Visser describes such a technique and explains its benefits in certain environments. (*Id.*)

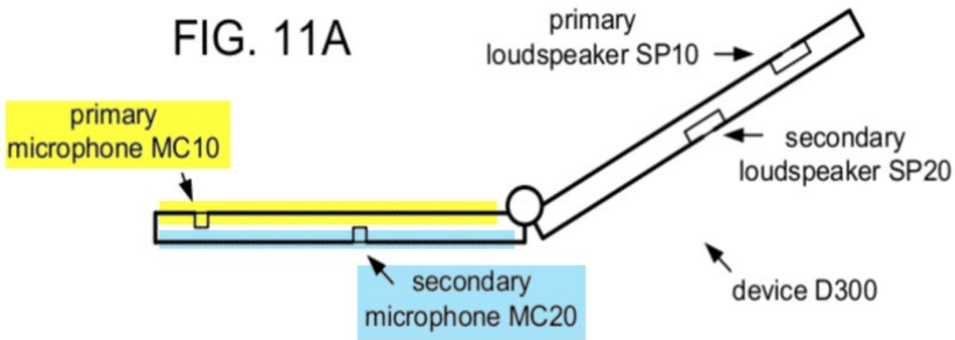
Fifth, determining DOA based on both phase differences and magnitude differences would represent the use of a known technique (e.g., Visser's use of both techniques) to improve similar devices (e.g., the Chen-Li DOA unit) in the same way (e.g., improved DOA determination). (*Id.* ¶143); *KSR*, 550 U.S. at 417. 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). (*Id.*)

A POSITA would have reasonably expected 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 similar to Chen's (e.g., a phone). (EX-1002 ¶144.) Thus, implementing the combined DOA approach would have been within a POSITA's skill. (*Id.* ¶144.)

D. Claim 12

Claim 12 also would have been obvious in further view of Visser. Visser discloses a mobile phone having a primary microphone MC10 and a secondary microphone MC20 on opposite faces:

FIG. 11A



(EX-1012, Fig. 11A, 10:31-35; EX-1002 ¶147.)

A POSITA would have been motivated to implement Visser's microphone orientation in Chen's phone. First, Chen teaches that microphones can be placed anywhere on a phone. (*Supra* §VII.I.) A POSITA would have been motivated to look at references, like Visser, to determine arrangements for reducing noise. (EX-1002 ¶148.)

Second, Chen teaches orienting one microphone towards an interviewee and another towards the interviewer. (EX-1003 ¶[0018].) Visser's microphone placement would provide this desired orientation. (EX-1002 ¶149.)

Third, in a multi-microphone system for noise reduction, a POSITA 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). (*Id.* ¶150.)

A POSITA would have expected success in implementing Visser's microphone arrangement because microphone placement is a simple design choice, as evidenced by Chen's disclosure that they can be located anywhere on the device. (*Id.* ¶151.)

Thus, claim 12 would have been obvious over the references in Grounds 1A or 1B and Visser. (*Id.* ¶¶146-52.)

E. Claim 20

Claim 20 depends from claim 19 and adds a similar limitation to claim 9. Claim 20 would have been obvious for the reasons discussed for claims 19 and 9. (*Supra* §§VII.G-VII.H, IX.C; EX-1002 ¶¶153-54.)

X. GROUND 1D: CLAIMS 10 AND 11 WOULD HAVE BEEN OBVIOUS IN VIEW OF THE REFERENCES IN GROUND 1A OR 1B, KIM, AND KLEFFNER.

Claims 10 and 11 each depend from claim 1. The additional limitations in Claims 10 and 11 are disclosed by or would have been obvious over Kim and Kleffner. Thus, claims 10 and 11 would have been obvious over the references in Ground 1A (Chen-Li) or Ground 1B (Chen-Li-Florencio), as applied to claim 1, in further view of Kim and Kleffner. (EX-1002 ¶155.)

Petitioners incorporate the discussions of Chen, Li, and Florencio from Grounds 1A and 1B.

A. Claim 10

Claim 10 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."

Li discloses using a sampling rate of 8 kHz for its adaptive filter. (EX-1004 ¶[0052].) Thus, Li teaches the second part of claim 10, i.e., supplying signals to the adaptive filter at a frequency of 12 kHz or lower. (EX-1002 ¶157.) Neither Chen nor Florencio discuss sampling frequencies for their VAD or DOA determination. (*Id.*)

Kim discloses processing an audio signal using various sampling frequencies, including 48 kHz. (EX-1006, 25, 14.) Kim estimates the DOA of an audio signal before downsampling the audio signal "in order to maximize the angle resolution of the DOA." (*Id.*[2])

Kleffner discloses audio signals with multiple sampling frequencies. (EX-1007, 26.) For example, Kleffner discloses audio signals sampled at 32 or 48 kHz downsampled to 8 kHz. (*Id.*; EX-1002 ¶159.)

A POSITA would have been motivated to use an audio signal with a higher sampling frequency (e.g., 48kHz) in Chen's DSP unit to determine the DOA. (EX-1002 ¶¶160-67.) First, although Li teaches using an 8 kHz sampling frequency

[2] Downsampling reduces the number of samples in an audio signal. (EX-1002 ¶158.) 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). (*Id.*)

for the noise reduction stage, Chen does not disclose the sampling frequency used for determining DOA. (EX-1002 ¶161) Thus, a POSITA would have been motivated to look to references such as Kim and Kleffner. (*Id.*)

Second, Kim discloses processing audio signals with a higher sampling frequency to determine the DOA and then downsampling the signal for further processing. (EX-1006, 25.) Kim explains that determining the DOA with a higher sampling frequency maximizes the angle of arrival. (*Id.*; *id.*, 19 ("a higher sampling rate is better" for DOA); EX-1002 ¶162.)

Third, a POSITA would have been motivated to use a higher sampling frequency, e.g., 48 kHz, because it was a standard sampling frequency for music and other audio signals. (EX-1002 ¶163.) It was also taught by both Kim and Kleffner. (EX-1007, 26.)

Fourth, using a higher frequency such as 48 kHz would use a known technique (e.g., Kim's higher frequency) to improve similar devices (e.g., Chen's DOA) in the same way (e.g., improved DOA accuracy). (EX-1002 ¶164); *KSR*, 550 U.S. at 417. Moreover, it would apply a known technique (e.g., Kim's higher 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). (*Id.*)

Fifth, while using a high sampling frequency improves the DOA determination, a POSITA would have understood it is computationally intensive to use a high sampling frequency at every stage. (EX-1002 ¶165.) Thus, a POSITA would have been motivated to downsample the audio signal for the remainder of the processing to reduce processing requirements in the Chen-Li (and Florencio, for Ground 1B) combination. (*Id.*) Kim and Kleffner both disclose downsampling the audio signal after using the higher sampling frequency. (EX-1006, 25; EX-1007, 26.) And, Li teaches that 8kHz is sufficient for noise reduction. (EX-1002 ¶165.) Thus, a POSITA would have been motivated to downsample the audio signal to a lower sampling frequency (e.g., 8kHz) for the noise reduction stage. (*Id.*)

A POSITA would have reasonably expected success in using higher sampling frequencies (e.g., 48 kHz) in Chen's system because they reflect standard, widely known sampling frequencies. (*Id.* ¶166.) There are a finite number of sampling frequencies audio processing devices use. (*Id.*) Moreover, downsampling audio signals to reduce computational requirements was well-known. (*Id.*) 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. (*Id.*)

B. Claim 11

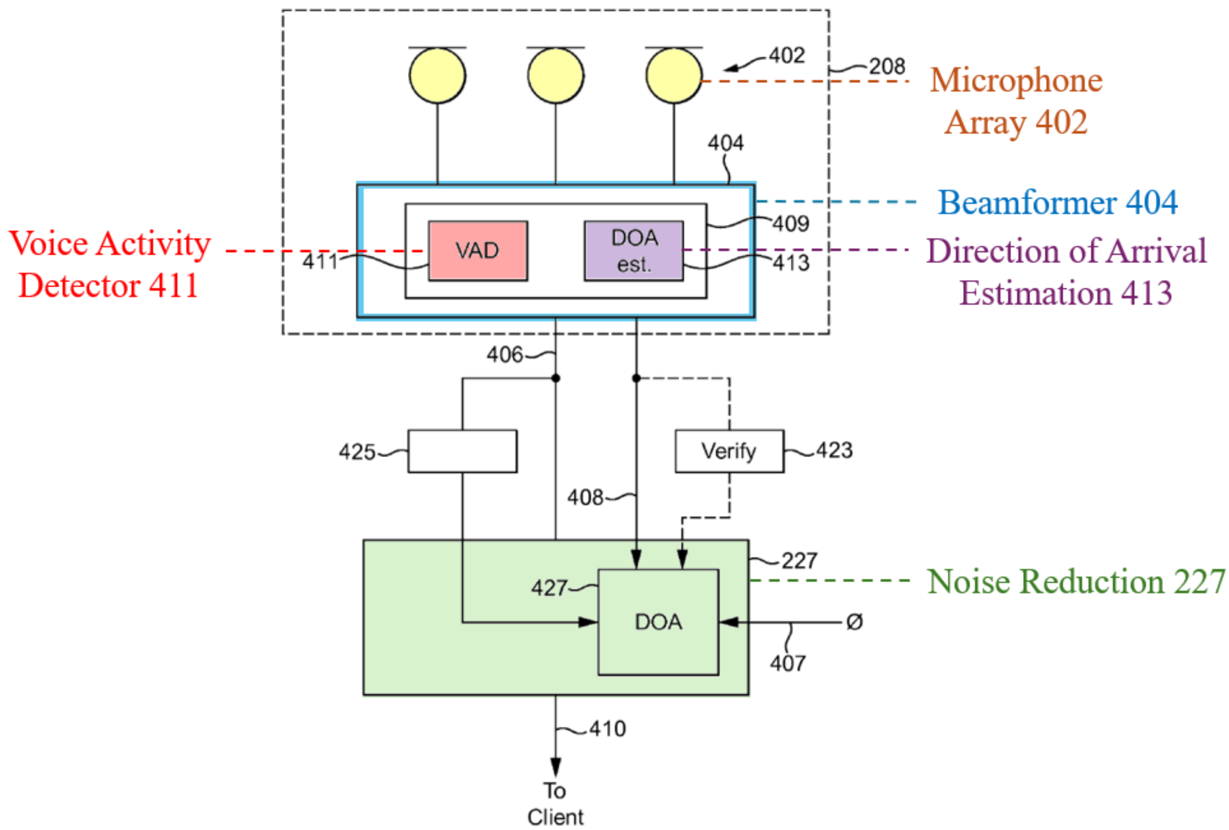
Claim 11, which depends from claim 1, recites "the speech segment determiner outputs more accurate speech segment information to the voice direction detector than speech segment information to the adaptive filter."

As explained above for claim 10, a POSITA would have been motivated to use a signal with a higher sampling frequency in the DOA unit and a lower sampling frequency in the noise reduction stage. 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). Thus, this limitation would have been obvious over the references in Ground 1A (Chen-Li) or Ground 1B (Chen-Li-Florencio) and further in view of Kim and Kleffner for the same reasons as claim 10. (EX-1002 ¶¶168-70.)

XI. GROUND 2A: CLAIMS 1-3, 9, AND 13-15 WOULD HAVE BEEN OBVIOUS IN VIEW OF STRÖMMER.

Strömmer discloses a system for processing audio signals in a device (e.g., mobile phone) having a microphone array and a beamformer. (EX-1010, Abstract, ¶¶[0041], [0053].)

Strömmer's beamformer contains a VAD, which determines whether a signal contains speech by detecting "speech like qualities" in the signal. (*Id.*) The beamformer also contains a 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 also includes a subsequent noise reduction stage 227 that further reduces noise:



(EX-1010, Fig. 4, ¶¶[0056]-[0059]; EX-1002 ¶¶171-72.)

A. Claim 1

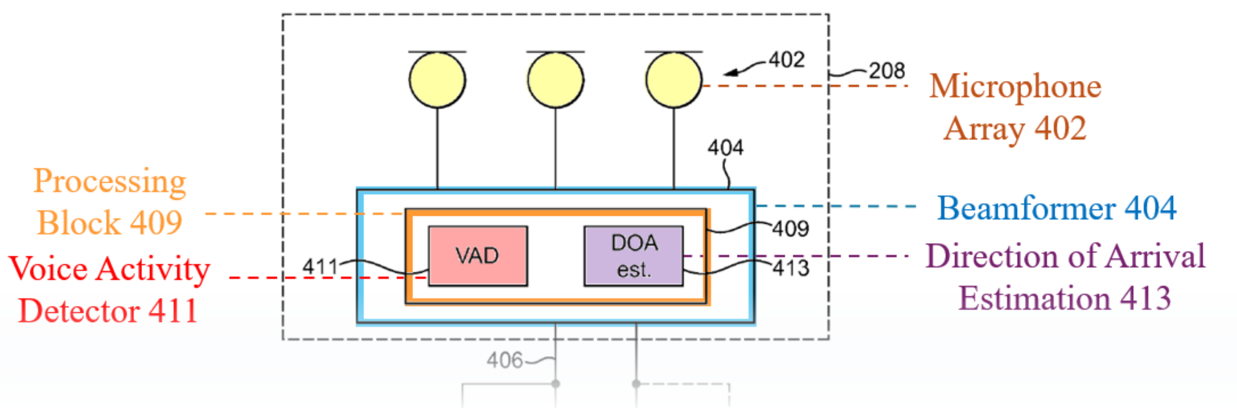
1. Preamble[3]

Strömmer's device reduces noise in audio signals using a beamformer and a "noise reduction stage." (EX-1010, Abstract, ¶¶[0053], [0064]-[0065], [0057], Fig. 4.) Accordingly, Strömmer discloses the preamble. (EX-1002 ¶173.)

[3] Because the claim language was provided in Ground 1, it is not repeated.

2. Element 1[a]

Strömmer discloses a microphone array 402 and a beamformer 404. (EX-1010 ¶¶[0042], [0052].) As shown below, the beamformer's processing block 409 comprises a 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.)



(EX-1010, Fig. 4 (excerpt); EX-1002 ¶175.)

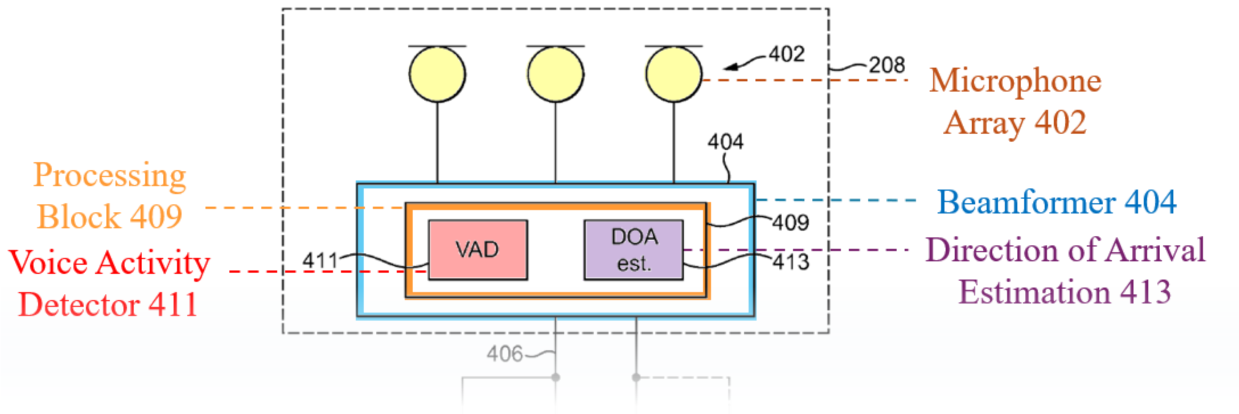
Accordingly, Strömmer discloses a speech segment determiner (e.g., VAD) that determines whether a sound picked up by at least either a first microphone or a second microphone (e.g., microphones in array 402) is a speech segment as recited in element 1[a][i]. (EX-1002 ¶176.)

Strömmer's VAD outputs speech segment information to processing block 409 and beamformer 404. (EX-1002 ¶177.) Strömmer's processing block 409 "ascertains the nature of the audio signals" and determines the DOA "based on detection of speech like qualities detected by the VAD 11[.]" (EX-1010 ¶[0053]; *id.* ¶¶[0067], [0047].) The beamformer 404 then forms a beam with a high gain in the direction of the wanted speaker signal. (*Id.*)

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. (EX-1002 ¶¶174-78.)

3. Element 1[b]

Strömmer's processing block 409 contains a DOA estimation block 413:



(EX-1010, Fig. 4 (excerpt).) The block 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 microphones. (*Id.* ¶¶[0061], [0053], [0062] ("direction of arrival … estimated" by block 413); EX-1002 ¶180.)

Based on speech detection by VAD 11 and DOA information estimated in block 413, processing block 409 determines the "principal direction(s) of main speaker(s)[.]" (*Id.* ¶[0053]; *id.* ¶¶[0047] (desired signals identified by "detection of speech like qualities and a principal direction of a main speaker is determined"), [0067]; EX-1002 ¶181.)

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., principal direction of main speaker), 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., signals obtained by first and second microphones in array 402) as recited in element 1[b][i]. (EX-1002 ¶182.)

Strömmer also discloses, in two ways, 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]. (*Id.* ¶183.) First, Strömmer's block 409 outputs "DOA information" to beamformer 404, which "uses the DOA information to process the audio signal by forming a beam." (EX-1010 ¶[0053]; *id.* ¶[0067].) Second, Strömmer discloses "DOA information estimated in the beamformer 404 is supplied to the noise reduction stage 227 and to signal processing circuitry 420." (*Id.* ¶[0055].)

Thus, 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. (EX-1002 ¶¶179-84.)

Accordingly, Strömmer discloses or renders obvious element 1[b][ii]. (*Id.*)

4. Element 1[c]

Strömmer discloses two adaptive filters: (a) beamformer 404 and (b) noise reduction stage 227. (EX-1002 ¶185.)

a. Beamformer

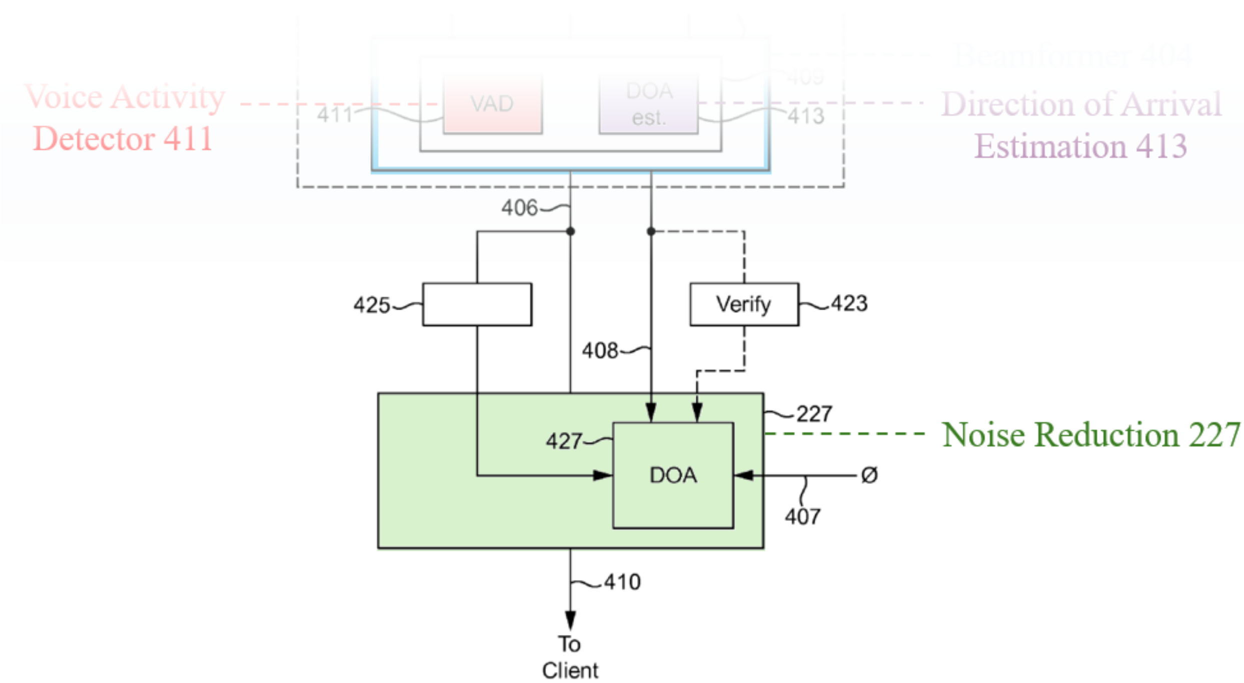
A beamformer is a spatial filter that reduces noise by suppressing the signal from all directions except the desired ones. (EX-1002 ¶186; EX-1011, 3.) An "adaptive" beamformer is an "adaptive filter." (EX-1002 ¶186.) Indeed, Strömmer teaches that adaptive beamformers "use the DOA information to filter the signals from the microphones in an array[.]" (EX-1010 ¶[0008].)

Strömmer's adaptive 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." (EX-1010 ¶[0053].) As explained, Strömmer's adaptive beamformer uses (via processing block 409) information from the VAD regarding whether the signal contains speech and DOA information from block 413 to determine the DOA for wanted signals (e.g., speech). (*Supra* §§XI.A.2-XI.A.3; EX-1002 ¶187.)

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., 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). (EX-1002 ¶¶186-88.)

b. Noise Reduction

Strömmer also discloses a noise reduction stage 227:



(EX-1010, 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].) The beamformer audio output contains a filtered combination of the microphone pickup signals. (*Id.* ¶[0008]; EX-1016, 946-53; EX-1002 ¶189.)

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 either "determines a level of noise suppression using conventional methods" or applies "maximum attenuation." (EX-1010 ¶¶[0068], [0071]-[0072]; EX-1002 ¶190.) The noise reduction stage is an adaptive filter because it adaptively changes the attenuation level based on current DOA information. (EX-1002 ¶190.)

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 filtered combination of signals from first/second microphones) based on the speech segment information (e.g., whether signal contains speech-like qualities) and the voice incoming-direction information (e.g., DOA information). (*Id.* ¶¶189-91.)

Thus, Strömmer discloses or renders obvious every limitation of claim 1, and the claim as a whole would have been obvious. (*Id.* ¶¶173-92.)

B. Claim 13

Independent claim 13 is unpatentable for the same reasons as claim 1. (*Id.* ¶193.)

C. Claims 2 and 14

Strömmer discloses using the time difference estimation (e.g., phase difference) to determine the DOA. (EX-1010 ¶[0061]; *see also* EX-1001, 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 ¶[0062].) Strömmer illustrates this delay in Figure 5, which shows two microphones 403 and 405 receiving audio signals from a source 516:

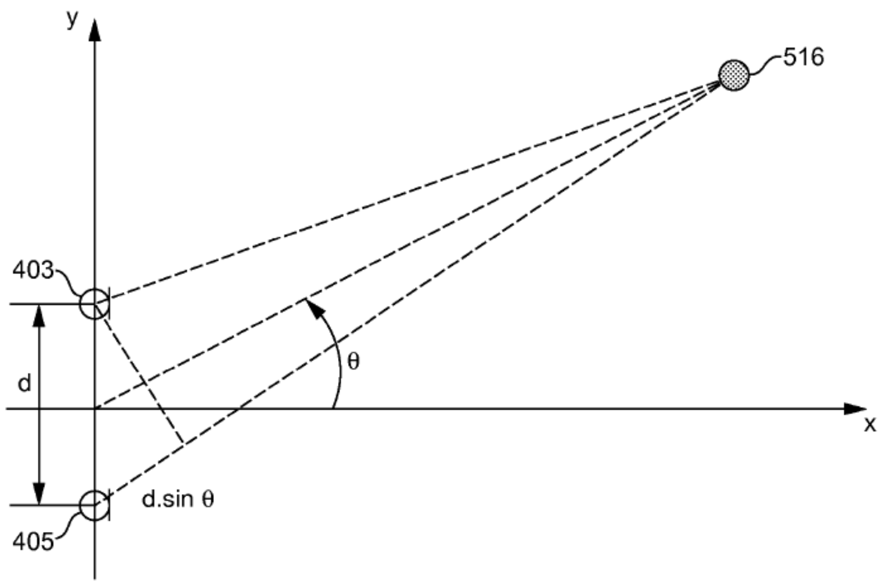


FIG. 5

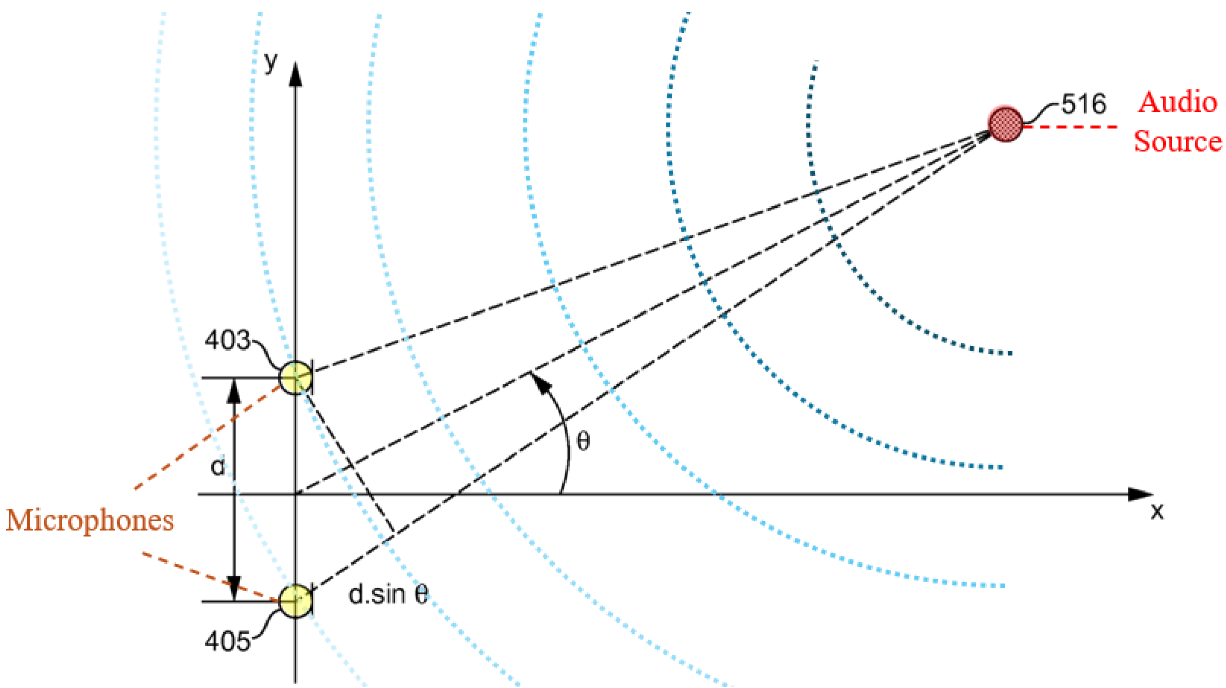
(*Id.*, Fig. 5.)

Strömmer's system estimates time delays "using correlation methods" and the known microphone locations. (*Id.* ¶[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 POSITA 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-1002 ¶¶195-96; EX-1016, 1046-47.)

Accordingly, 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. (EX-1002 ¶¶194-98.)

D. Claims 3 and 15

As explained for claims 2 and 14, Strömmer discloses using phase differences (e.g., time delay) to determine the DOA as illustrated in Figure 5:



(EX-1010, Fig. 5 (annotated).) As shown by the dotted blue lines (as added by Petitioner's expert), the sound waves from a source arrive at the microphones (403,

405) at different times. (EX-1002 ¶200.) A POSITA would have understood that the microphone that detects the sound first will be the closest source. (*Id.*) In other words, the sound signal detected by the first microphone will have a more advanced phase if the source is closer to that microphone. (*Id.*) In Figure 5, the audio signal detected by microphone 403 has a more advanced phase than the signal detected by microphone 405. (*Id.*) If the audio source were below the x-axis in Figure 5, the signal detected by microphone 405 would have a more advanced phase. (*Id.*)

Strömmer's beamformer 404 forms a beam with "a high gain" in the speaker's direction and "a low gain" in other directions. (EX-1010 ¶[0053].) To form the beam, Strömmer applies coefficients to the microphone array signals. (EX-1002 ¶201; *see* EX-1011, 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. (*Id.*)

Thus, a POSITA 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. (EX-1002 ¶¶199-203.)

E. Claim 9

Strömmer's VAD detects speech in each received sound signal. (EX-1010 ¶[0053].) Thus, a POSITA 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 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). (EX-1002 ¶¶204-06.)

XII. GROUND 2B: CLAIMS 4, 12, AND 16 WOULD HAVE BEEN OBVIOUS IN VIEW OF STRÖMMER AND VISSER.

A. Claims 4 and 16

Visser discloses or renders obvious the additional limitation of claims 4 and 16. (*Supra* §IX.A.) A POSITA would have incorporated Visser's processing of broadside signals into Strömmer's device, and would have reasonably expected success in doing so, for the same reasons a POSITA would have done so for Chen.

(*Supra* §IX.A; EX-1002 ¶209.) Thus, claims 4 and 16 would have been obvious over Strömmer and Visser. (EX-1002 ¶¶208-10.)

B. Claim 12

Visser discloses the additional limitation recited in claim 12. (*Supra* §IX.D.) A POSITA 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 POSITA would have been motivated to do so for Chen. (*Supra* §IX.D; EX-1002 ¶212.) Thus, claim 12 would have been obvious over Strömmer and Visser. (EX-1002 ¶¶211-13.)

XIII. GROUND 2C: CLAIMS 5-6 AND 17-18 WOULD HAVE BEEN OBVIOUS IN VIEW OF STRÖMMER AND BRANDSTEIN.

Claims 5-6 and 17-18 would have been obvious over Strömmer (as in claims 1 and 13) in view of Brandstein. (EX-1002 ¶214.) Petitioners incorporate the discussion of Strömmer from Ground 2A here and discuss the claim elements for which Brandstein is relevant below.

Brandstein is a textbook on digital signal processing of microphone arrays. (EX-1011.) Brandstein discloses multiple techniques for SSL and speech recognition. (*E.g.*, *id.*, 157-201 (SSL), 331-53 (speech recognition); EX-1002 ¶215.)

A. Claims 5 and 17

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

In SRP-based SSL, a beamformer is steered "to specific spatial points of interest" to evaluate the output signal, typically the power. (*Id.*, 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. (*Id.*) The beamformer output is a function of the microphone signals, as shown in Equation 8.15. (*Id.*) A POSITA would have understood based on these equations that SRP is a function of the microphone signals' magnitudes. (EX-1002 ¶¶217-18.)

Brandstein also discloses SRP SSL in the frequency domain using a phase transform (SRP-PHAT). (EX-1011, 170-72.) SRP-PHAT is "similar to the standard SRP-based approaches." (*Id.*, 171.) SRP-PHAT looks for maximum power over a region of potential source locations. (*Id.*, 172.) But SRP-PHAT includes a "PHAT weighting" that provides "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 DOA using microphone signal magnitudes—SRP and SRP-PHAT. (EX-1002 ¶219; EX-1012, 28:60-64 (SRP-PHAT is based on gain differences).)

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). (EX-1002 ¶¶216-20.)

A POSITA would have been motivated to use SRP or SRP-PHAT SSL in Strömmer's system for several reasons. (*Id.* ¶¶221-27.)

First, Brandstein explains that SRP and SRP-PHAT offer advantages in certain environments. SRP is "optimal" for environments that have a "limited case of additive, uncorrelated, and uniform variance noise and equal source-microphone distances." (EX-1011, 169.) Thus, a POSITA would have been motivated to use SRP in Strömmer's system when the environment meets these criteria. (EX-1002 ¶222.)

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, 171.) SRP-PHAT provides "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 localization methods[.]" (*Id.*, 172.) These advantages would have motivated a POSITA to replace the cross-correlation SSL in Strömmer with the SRP-PHAT SSL in Brandstein. (EX-1002 ¶223.) Indeed, Brandstein describes experiments showing SRP-PHAT outperformed "TDOA-based localization methods," which are the type of methods Strömmer uses. (EX-1011, 177; *id.*, 161 (TDOA used cross-correlation); EX-1010 ¶¶[0061]-[0063] (relying on cross-correlation).)

Third, SRP and SRP-PHAT are known SSL techniques (EX-1011, 158) and using them in Strömmer would yield predictable results (e.g., more accurate SSL). (EX-1002 ¶224.) It would also reflect the 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). (*Id.*)

Further, a POSITA would have found it obvious to try SRP and/or SRP-PHAT because only a finite number of SSL techniques existed. (EX-1011, 158; EX-1002 ¶225.)

A POSITA would have reasonably expected success in using SRP or SPR-PHAT techniques in Strömmer's system. (EX-1002 ¶226.) Brandstein describes multiple SSL methods, including the cross-correlation methods. (EX-1011, 161-

62.) Thus, a POSITA 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. (EX-1002 ¶226.)

B. Claims 6 and 18

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 ¶[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. (EX-1002 ¶229; EX-1011, 8-11, 88-90.)

The cancellation of noise components in a beamformer is mutual. Noise components in a first signal cancel the noise components in a second signal. (EX-1002 ¶230.) At the same time, noise components in the second signal cancel the noise components in the first signal. (*Id.*) This mutual cancellation occurs regardless of which microphone signal has a greater magnitude. (*Id.*)

As explained 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. (*Id.* ¶¶228-32.)

XIV. GROUND 2D: CLAIMS 7-8 AND 19-20 WOULD HAVE BEEN OBVIOUS IN VIEW OF STRÖMMER, BRANDSTEIN, AND VISSER.

A. Claim 7

Visser discloses or renders obvious claim 7's additional limitation. (*Supra* §IX.B.) A POSITA would have been motivated to incorporate Visser's SSL and noise reduction for broadside sources into Strömmer's device, and would have reasonably expected success in doing so, for the same reasons discussed for Chen. (*Supra* §§IX.A-IX.B; EX-1002 ¶¶234-36.)

B. Claims 8 and 19

Strömmer discloses using phase difference and Brandstein discloses using magnitude. (*Supra* §§XI.C, XIII.A.) Visser discloses or renders obvious the additional limitation of claims 8 and 19. (*Supra* §IX.C.) Thus, Strömmer, Brandstein, and Visser together render obvious detecting the voice incoming 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). (EX-1002 ¶¶237-38.)

A POSITA would have combined multiple SSL techniques in Strömmer's device, as taught by Visser, and would have reasonably expected success in doing so, for the same reasons discussed with Chen. (*Supra* §IX.C; EX-1002 ¶¶239-40.)

C. Claim 20

Claim 20 is unpatentable for the same reasons as claim 9. (*Supra* §XI.E; EX-1002 ¶¶241-42.)

XV. GROUND 2E: CLAIMS 10 AND 11 WOULD HAVE BEEN OBVIOUS IN VIEW OF STRÖMMER, KIM, AND KLEFFNER.

Claims 10 and 11 would have been obvious over Strömmer (as discussed for claim 1) in view of Kim and Kleffner. (EX-1002 ¶243.)

A. Claim 10

As discussed in Ground 1D, Kim and Kleffner together disclose and render obvious this limitation. (*Supra* §X; EX-1002 ¶244.)

A POSITA 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. (EX-1002 ¶¶245-48.) First, Strömmer does not describe what sampling frequency to use. (*Id.* ¶246.) Thus, a POSITA would have been motivated to look to references such as Kim and Kleffner to see what sampling frequency to use. (*Id.*) Moreover, a POSITA 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 as for incorporating the technique into Chen in Ground 1D. (*Supra* §X.A; EX-1002 ¶247.)

B. Claim 11

As explained for claim 10, a POSITA would have been motivated to use a signal with a higher sampling frequency in Strömmer's DOA unit. A POSITA 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). (EX-1002 ¶250.) Thus, a POSITA 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 reasons discussed for claim 10, a POSITA 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). (*Id.* ¶¶249-51.)

XVI. GROUND 3A: CLAIMS 1-3, 9, AND 13-15 WOULD HAVE BEEN OBVIOUS IN VIEW OF FLORENCIO.

As discussed, Florencio's system includes a microphone array, a VAD (306), an SSL unit (308), and an adaptive beamformer (310):

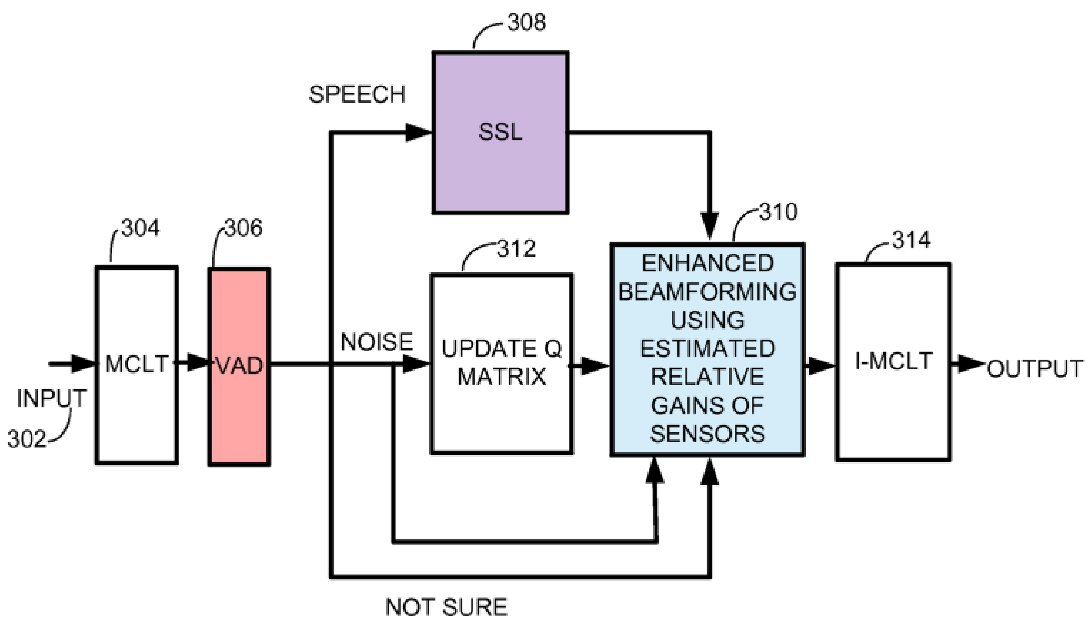


FIG. 3

(EX-1005, Fig. 3, ¶¶[0020]-[0024], [0044]-[0045], [0048], Figs. 1-7; EX-1002 ¶252.)

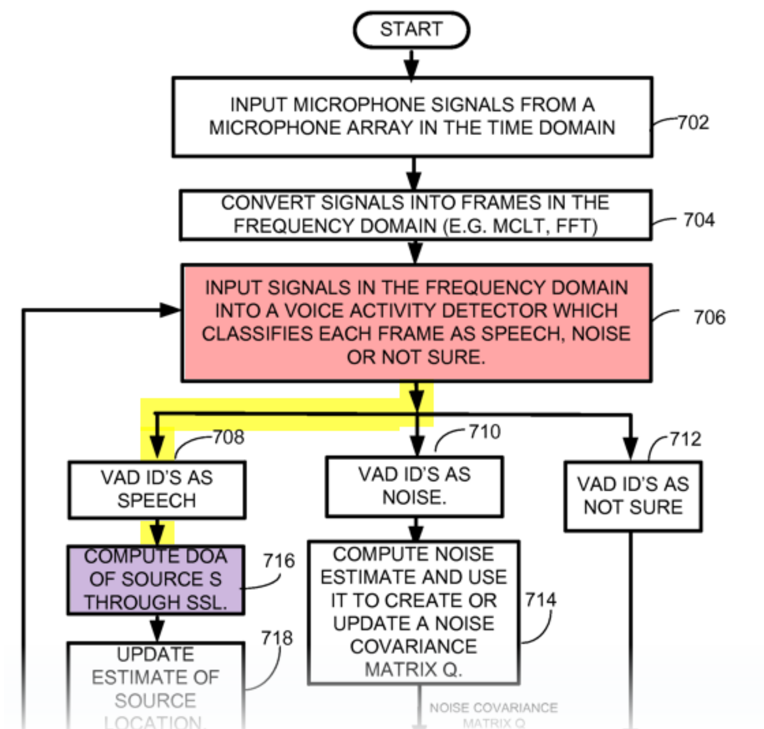
A. Claim 1

1. Preamble

Florencio discloses noise reduction using an "enhanced beamforming technique" that can be implemented on "a general purpose computing device." (EX-1005 ¶¶[0012], [0002], Fig. 1; EX-1002 ¶253.) Accordingly, Florencio discloses the preamble. (EX-1002 ¶253.)

2. Element 1[a]

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 ¶¶[0044]-[0045], [0048], Figs. 3-4, 7.) If the VAD classifies the frame as Speech, then SSL occurs to better estimate the signal's DOA. (*Id.*)

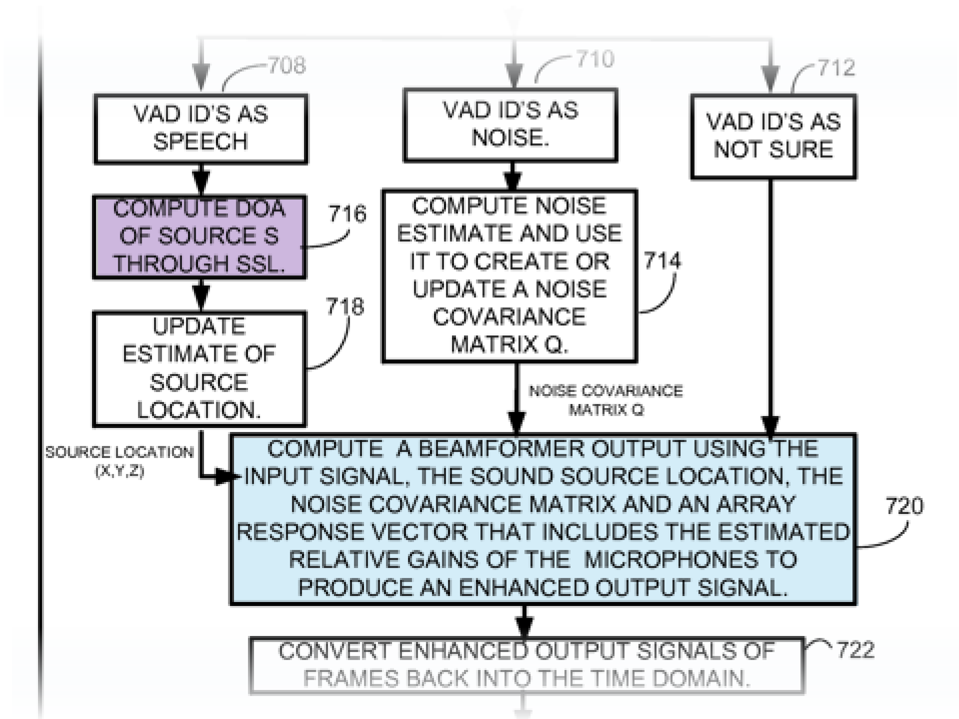


(*Id.*, Fig. 7 (excerpt); EX-1002 ¶255.)

Accordingly, 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. (EX-1002 ¶¶254-56.)

3. Element 1[b]

Florencio discloses that the VAD output (e.g., whether the sound contains speech) is provided to the SSL module. (EX-1005 ¶¶[0044], [0048], Fig. 7, claim 9; EX-1002 ¶258.) When the sound contains speech, the SSL module "obtain[s] a better estimate of the location of the desired signal" (EX-1005 ¶[0044], claim 13) as 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].)

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., signals obtained by first/second microphones) and to output (e.g., to the beamformer) voice incoming-direction information when the voice incoming direction is detected. (EX-1002 ¶¶257-59.)

4. Element 1[c]

Florencio explains that adaptive beamformers (i.e., filters) were known (EX-1005 ¶¶[0001]-[0004]) and discloses an "enhanced" adaptive beamformer. (*Id.* ¶¶[0006], [0025], [0031], [0037]-[0043], [0048]; EX-1002 ¶261.)

In addition to the source location and speech frames (e.g., speech signals from the microphone array) (*supra* §XVI.A.3), Florencio discloses using the sound data classified as "Noise" or "Not Sure" in the beamformer. (EX-1005 ¶¶[0044], [0046]-[0048], Figs. 3, 5-7; EX-1002 ¶262.) The beamformer is adaptive, e.g., parameters like gain are modified based on changing conditions. (*Id.*)

Florencio's adaptive beamformer performs a noise reduction process. (EX-1005 ¶¶[0007], [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], [0045], [0047], [0048] (received signals provided to beamformer), Figs. 4-6, claims 1, 6, 9-10, 13; EX-1002 ¶263.)

Florencio's adaptive beamformer performs noise reduction based on the sound source location information. (EX-1005 ¶¶[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; EX-1002 ¶264.)

Florencio therefore discloses or renders obvious an adaptive filter (e.g., adaptive beamformer) configured to perform a noise reduction process (e.g., increase 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). (EX-1002 ¶¶260-66.)

B. Claim 13

Independent claim 13 is unpatentable for the same reasons as claim 1. (*Id.* ¶267.)

C. Claims 2 and 14

Florencio discloses using an SSL algorithm "based on time delay of arrival of the signal and maximum likelihood estimation." (EX-1005 ¶[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.*; EX-1002 ¶¶268-70.)

D. Claims 3 and 15

Florencio discloses using a phase difference. (*Supra* §XVI.C.) As explained above, a POSITA 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. (*Supra* §XI.D; EX-1002 ¶272.) 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. (*Supra* §XVI.A.4; EX-1002 ¶272.) Florencio therefore discloses claims 3 and 15. (EX-1002 ¶¶271-73.)

E. Claim 9

In an array, the microphone closest to the speaker will detect a sound signal with a more advanced phase. (EX-1002 ¶275.) Florencio analyzes each microphone signal for speech. (EX-1005 ¶[0048], Fig. 7.) Thus, a POSITA 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). (EX-1002 ¶¶274-76.)

XVII. GROUND 3B: CLAIMS 4, 12, AND 16 WOULD HAVE BEEN OBVIOUS IN VIEW OF FLORENCIO AND VISSER.

A. Claims 4 and 16

Visser discloses or renders obvious the additional limitation of claims 4 and 16. (*Supra* §IX.A.) A POSITA would have incorporated Visser's SSL and noise reduction into Florencio's device, and would have reasonably expected success in doing so, for the same reasons described for Chen and Strömmer. (*Supra* §§IX.A, XII.A; EX-1002 ¶279.) Thus, claims 4 and 16 would have been obvious over Florencio and Visser. (EX-1002 ¶¶278-80.)

B. Claim 12

Visser discloses or renders obvious the additional limitations of claim 12. (*Supra* §IX.D.) A POSITA would have incorporated Visser's microphone arrangement into Florencio's device, and would have reasonably expected success in doing so, for the same reasons as discussed for Chen and Strömmer. (*Supra* §§IX.D, XII.B; EX-1002 ¶282.) Thus, claim 12 would have been obvious over Florencio and Visser. (EX-1002 ¶¶281-83.)

XVIII. GROUND 3C: CLAIMS 5-6 AND 17-18 WOULD HAVE BEEN OBVIOUS IN VIEW OF FLORENCIO AND BRANDSTEIN.

Claims 5-6 and 17-18 would have been obvious over Florencio in view of Brandstein. (EX-1002 ¶284.) Petitioners incorporate the discussion of Florencio from Ground 3A here.

A. Claims 5 and 17

Brandstein discloses or renders obvious the additional limitation of claims 5 and 17. (*Supra* §XIII.A.) A POSITA would have incorporated Brandstein's SRP or SRP-PHAT into Florencio's device, and would have reasonably expected success in doing so, for the same reasons as for incorporating SRP or SRP-PHAT into Strömmer. (*Supra* §XIII.A; EX-1002 ¶286.) Thus, claims 5 and 17 would have been obvious over Florencio and Brandstein. (EX-1002 ¶¶285-87.)

B. Claims 6 and 18

Florencio's beamformer, as modified to include Brandstein's SRP or SRP-PHAT (as discussed for claim 5 immediately above), discloses or renders obvious the additional limitation of claims 6 and 18 for the same reasons as the Strömmer-Brandstein combination . (*Supra* §§XIII.B, XVIII.A.) Thus, claims 6 and 18 would have been obvious over Florencio and Brandstein. (EX-1002 ¶¶288-89.)

XIX. GROUND 3D: CLAIMS 7-8 AND 19-20 WOULD HAVE BEEN OBVIOUS IN VIEW OF FLORENCIO, BRANDSTEIN, AND VISSER.

A. Claim 7

Visser discloses or renders obvious the additional limitation of claim 7. (*Supra* §IX.B.) A POSITA would have incorporated Visser's SSL and noise reduction into Florencio's device, and would have reasonably expected success in doing so, for the same reasons as for Chen and Strömmer. (*Supra* §§IX.A-IX.B, XIV.A; EX-1002 ¶292.) Thus, claim 7 would have been obvious over Florencio, Brandstein, and Visser. (EX-1002 ¶¶291-93.)

B. Claims 8 and 19

Visser discloses or renders obvious the additional limitation of claims 8 and 19 (*supra* §IX.C), and a POSITA would have used multiple SSL techniques in Florencio's device, as taught by Visser, and would have reasonably expected success in doing so, for the same reasons discussed for Chen and Strömmer. (*Supra* §§IX.C, XIV.B; EX-1002 ¶¶294-96.)

C. Claim 20

Claim 20 is unpatentable for the same reasons as claim 9. (*Supra* §XVI.E; EX-1002 ¶¶297-98.)

XX. GROUND 3E: CLAIMS 10 AND 11 WOULD HAVE BEEN OBVIOUS IN VIEW OF FLORENCIO, KIM, AND KLEFFNER.

Kim and Kleffner disclose the limitations of claims 10 and 11. (*Supra* §X; EX-1002 ¶300.) A POSITA 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, for the same reasons a POSITA would have been motivated to do so in Chen's and Strömmer's systems. (*Supra* §§X, XV; EX-1002 ¶301.) A POSITA also would have reasonably expected success for the same reasons. (*Id.*) Thus, these claims would have been obvious over Florencio, Kim, and Kleffner. (EX-1002 ¶¶299-302.)

XXI. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

Where, as here, a strong prima facie obviousness showing exists, secondary considerations may not dislodge the obviousness conclusion. *Leapfrog Enters., Inc. v. Fisher-Price, Inc.*, 485 F.3d 1157, 1162 (Fed. Cir. 2007). Petitioners are aware of no evidence supporting a claim for secondary considerations.

XXII. CONCLUSION

Petitioners request the Board institute trial and cancel all challenged claims.[4]

[4] Petitioners will address discretionary denial issues if raised by PO. *See* Memorandum from Acting Director Stewart, *Interim Processes for PTAB Workload Management* (March 26, 2025).

XXIII. MANDATORY NOTICES, GROUNDS FOR STANDING, AND FEE PAYMENT

Pursuant to 37 C.F.R. §42.8(a)(1), the mandatory notices identified in 37 C.F.R. §42.8(b) are provided below as part of this Petition.

A. Real Party-In-Interest (37 C.F.R. §42.8(b)(1))

Amazon.com, Inc., Amazon.com Services LLC, and Amazon Web Services, Inc. are the real parties-in-interest.

B. Related Matters (37 C.F.R. §42.8(b)(2))

PO asserted the ’259 patent against Petitioners in district court. *SoundClear Technologies LLC v. Amazon.com, Inc. et al.*, No. 1:24-cv-00728 (E.D. Va.). After intradistrict transfer, that case is now No. 2:24-cv-00320.

If this IPR is instituted and the above proceeding is not stayed, Petitioners hereby stipulate not to pursue in that proceeding any ground of invalidity, against any claim challenged herein, that was raised or reasonably could have been raised in this Petition.

To the best knowledge of Petitioner, the ’259 patent is or has been involved in the following additional proceedings:

Name	Number	Court	Filed
SoundClear Technologies LLC v. Google LLC	2:24-cv-00321	E.D. Va.	May 1, 2024
Google LLC v. SoundClear Technologies LLC	IPR2025-00345	P.T.A.B.	Feb. 10, 2025

C. Lead and Backup Counsel (37 C.F.R. §42.8(b)(3))

Petitioner provides the following designation of counsel, all of whom are included in Customer No. 20,995 identified in Petitioner's Power of Attorney.

Lead Counsel	**Back-up Counsel**
Colin B. Heideman (Reg. No. 61,513) 2cbh@knobbe.com BoxSEAZN2L2103LP1@knobbe.com <u>Postal and Hand-Delivery Address:</u> Knobbe, Martens, Olson, & Bear, LLP 555 110th Ave. NE, Ste. 500 Bellevue, WA 98004 Telephone: (206) 405-2000 Facsimile: (206) 405-2001	Joseph R. Re (Reg. No. 31,291) 2jrr@knobbe.com Jeremy A. Anapol (Reg. No. 75,686) 2jaa@knobbe.com <u>Postal and Hand-Delivery Address:</u> Knobbe, Martens, Olson, & Bear, LLP 2040 Main Street, 14th Floor Irvine, CA 92614 Telephone: (949) 760-0404 Facsimile: (949) 760-9502 Christie R.W. Matthaei (Reg. No. 62,933) 2crw@knobbe.com Nathan D. Reeves (Reg. No. 77,806) 2ndr@knobbe.com Logan P. Young (Reg. No. 79,294) 2lpy@knobbe.com <u>Postal and Hand-Delivery Address</u>: Knobbe, Martens, Olson & Bear, LLP 555 110th Ave. NE, Ste. 500 Bellevue, WA 98004 Telephone: (206) 405-2000 Facsimile: (206) 405-2001

D. Service Information (37 C.F.R. §42.8(b)(4))

Please direct all correspondence to lead counsel and back-up counsel at the addresses shown above. Petitioner also consents to electronic service by email to BoxSEAZN2L2103LP1@knobbe.com.

E. Grounds for Standing (37 C.F.R. §42.104)

Petitioners certify that the '259 patent is available for IPR and are not barred or estopped from requesting IPR on the identified grounds. This Petition is being filed within one year of service of the original complaint against Petitioners in the district court litigation.

F. Payment of Fees (37 C.F.R. §42.103)

The Office may charge the §42.15(a) fee to Deposit Account No. 11-1410. Review of twenty claims is requested. Payment for any additional fees due may be charged to the above-referenced Deposit Account.

Respectfully submitted,

KNOBBE MARTENS OLSON & BEAR, LLP

Dated: June 2, 2025

/Colin B. Heideman /

Colin B. Heideman (Reg. No. 61,513)
Joseph R. Re (Reg. No. 31,291)
Jeremy A. Anapol (Reg. No. 75,686)
Christie R.W. Matthaei (Reg. No. 62,933)
Nathan D. Reeves (Reg. No. 77,806)
Logan P. Young (Reg. No. 79,294)

Counsel for Petitioners
Amazon.com, Inc.,
Amazon.com Services LLC, and
Amazon Web Services, Inc.

APPENDIX

Listing of Claims from U.S. 9,031,259	
Claim 1	
1[pre]	A noise reduction apparatus comprising:
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 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;
1[b]	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; and
1[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.
Claim 2	
--	The noise reduction apparatus according to claim 1, wherein the voice direction detector detects the voice incoming direction based on a phase difference between the first and second sound pick-up signals.

Listing of Claims from U.S. 9,031,259	
Claim 3	
--	The noise reduction apparatus according to claim 2, wherein 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 4	
--	The noise reduction apparatus according to claim 2, wherein 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 5	
--	The noise reduction apparatus according to claim 1, wherein the voice direction detector detects the voice incoming direction based on magnitudes of the first and second sound pick-up signals.
Claim 6	
--	The noise reduction apparatus according to claim 5, wherein 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.

Listing of Claims from U.S. 9,031,259	
Claim 7	
--	The noise reduction apparatus according to claim 5, wherein 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.
Claim 8	
--	The noise reduction apparatus according to claim 1, wherein 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 9	
--	The noise reduction apparatus according to claim 1, wherein 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 10	
--	The noise reduction apparatus according to claim 1, wherein 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	
--	The noise reduction apparatus according to claim 1, wherein the speech segment determiner outputs more accurate speech segment information to the voice direction detector than speech segment information to the adaptive filter.

Listing of Claims from U.S. 9,031,259	
Claim 12	
12[pre]	An audio input 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]	a speech segment determiner configured [i] to determine whether or not a sound picked up by at least either the first microphone or the 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;
12[d]	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; and
12[e]	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.
Claim 13	
13[pre]	A noise reduction method comprising the steps of:
13[a]	determining whether or not a sound picked up by at least either a first microphone or a second microphone is a speech segment;

Listing of Claims from U.S. 9,031,259	
13[b]	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 determined that the sound picked up by the first or the second microphone is the speech segment; and
13[c]	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.
Claim 14	
--	The noise reduction method according to claim 13, the voice incoming direction is detected based on a phase difference between the first and second sound pick-up signals.
Claim 15	
--	The noise reduction method according to claim 14, wherein 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.
Claim 16	
--	The noise reduction method according to claim 14, wherein when the phase difference is within a predetermined range, an adaptive filter outputs either the first or the second sound pick-up signal without performing the noise reduction process.

Listing of Claims from U.S. 9,031,259	
Claim 17	
--	The noise reduction method according to claim 14, wherein the voice incoming direction is detected based on magnitudes of the first and second sound pick-up signals.
Claim 18	
--	The noise reduction method according to claim 17, wherein 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.
Claim 19	
--	The noise reduction method according to claim 14, wherein 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.
Claim 20	
--	The noise reduction method according to claim 19, wherein the speech segment is detected 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 is detected 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.

CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. §42.24(d), the undersigned certifies that this **PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,031,259** contains 13,916 words according to the word-processing program used to prepare this paper. The foregoing word count complies with the 14,000 word type-volume limit specified by 37 C.F.R. §42.24(a)(1).

Dated: June 2, 2025

By: */Colin B. Heideman/*
Colin B. Heideman (Reg. No. 61,513)
KNOBBE MARTENS OLSON & BEAR, LLP

CERTIFICATE OF SERVICE

The undersigned hereby certifies that on the date below a copy of this **PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,031,259** and **ACCOMPANYING EXHIBITS** are being served on June 2, 2025 via Federal Express overnight mail on counsel of record for U.S. Patent No. 9,031,259 at the Correspondence Address of record below:

182086 – DAIGNAULT IYER LLP
8229 Boone Blvd., Suite 450
Vienna, VA 22182
United States

A courtesy copy is also being served via email on counsel for the patent holder in the pending litigation:

Daignault Iyer LLP
8229 Boone Blvd., Suite 450
Vienna, VA 22182
Email: DI-SoundClear@daignaultiyer.com

Dated: June 2, 2025

By: */Colin B. Heideman/*
Colin B. Heideman (Reg. No. 61,513)
KNOBBE MARTENS OLSON & BEAR, LLP