

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. IPR2025-00565
U.S. Patent No. 11,069,337

REPLY DECLARATION OF RICHARD STERN, PH.D.

TABLE OF CONTENTS

I.	CLAIM CONSTRUCTION -----	1
A.	“Classify” -----	1
B.	“Voice” -----	4
C.	Analyze a Voice to Classify it Based on Distance -----	5
II.	SHIN -----	7
A.	Shin Discloses Classifying Voice Based on Distance -----	7
B.	Shin Discloses Voice Analysis -----	10
III.	SHIMOMURA -----	12
A.	Shimomura Discloses Voice Classification -----	12
B.	Shimomura Discloses Voice Analysis -----	13
C.	Shimomura Renders the Claims Obvious -----	14

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

1. In preparing this Declaration, I considered the Declaration of Joseph McAlexander (EX-2013), the exhibits cited therein, and all other references identified herein.

2. I understand that my prior declaration was submitted as Exhibit 1002, and I refer to certain portions of that declaration herein. I remain unaware of any evidence that would suggest that the claims of the '337 patent are not anticipated and/or would not have been obvious.

3. I understand that Mr. McAlexander has applied a slightly different definition than I have for a person of ordinary skill in the art. (EX-2013 ¶33.) While I disagree with Mr. McAlexander as to that definition, my opinions would not change under his definition.

4. In this Declaration, I provide opinions and enumerate ways in which I disagree with Mr. McAlexander and his opinions as stated in his Declaration (EX-2013). However, this Declaration does not enumerate all ways in which I disagree with Mr. McAlexander, and the fact that I do not address some of Mr. McAlexander's statements should not be taken to mean that I agree with them.

I. CLAIM CONSTRUCTION

A. "Classify"

5. In Paragraph 37 of his declaration, Mr. McAlexander states that in one

embodiment, the voice classifying unit can “determine if a voice is a whisper or not and classify the voice accordingly.” (EX-2013 ¶37.) In Paragraph 38, he notes that “[o]ther embodiments of the voice classifying unit use the ... distance between the speaker and the device[.]” (*Id.* ¶38.) In Paragraph 52, he discusses that the specification describes using the “user-to-device distance” and a threshold “for the purpose of distinguishing one class of voice from another.” (*Id.* ¶52.) And in Paragraph 53, he states that the “claimed classifying unit uses those characteristics to classify voices” and that each characteristic—including distance—“can be used alternatively or in different combinations.” (*Id.* ¶53.) He notes in Paragraph 54 that a person of ordinary skill in the art may use only “one variable” to classify the voice, even if it is not “perfectly effective.” (*Id.* ¶54.) I agree with these statements and they confirm my opinion, stated in my prior declaration, that the voice classifying unit may classify the user’s voice based solely on the user-to-device distance. (EX-1002 ¶61.)

6. The ’337 patent also confirms my opinion that the classification can be based solely on the measured distance. The claims explicitly recite classifying the voice “based on the distance”—and nothing else—and the specification explicitly discloses classifying based on distance alone. (EX-1001 (’337 patent), claim 1 (“classify the voice as either a first voice or a second voice *based on the distance* between the user and the voice-content control device” (emphasis added)), 8:20-26 (“the distance can be used as a feature value to perform the classification to the first

voice V1A and the second voice V1B”), 8:30-34 (“the voice classifying unit 38 sets a threshold of the feature value, and classifies the voice V1 as either one of the first voice V1A or the second voice V1B based on whether the feature value exceeds the threshold”), 8:10-12 (classification can be performed “using any method”).)

7. In Paragraph 43, Mr. McAlexander states that “[a] POSITA would have understood ‘classify’ to require something discrete, not a mere measurement” and that “[a] POSITA would not have understood ‘classify’ to be used in the context of measuring a single, continuously changing variable (like volume or distance).” (EX-2013 ¶43.) I disagree with the premise of Mr. McAlexander’s statements. Measurements are discrete. The specification describes classifying the user’s voice based on measurable characteristics such as distance, volume, frequency, etc. (EX-1001 (’337 patent), 8:20-34.) And, the claims expressly require the classification to be based on a measurable characteristic—distance.

8. Also in Paragraph 43, Mr. McAlexander states that “[a] POSITA would have understood the claims of the ’337 patent to require a classification, or distinction between, at least two separately identifiable voices.” (EX-2013 ¶43.) To the extent Mr. McAlexander’s opinion is that the patented device must distinguish between two different people’s voices, I disagree. The claims recite receiving “a voice” from “a user”—which includes a *single* voice from a *single* user—and placing that voice into one of two separately identifiable *classes* (i.e., categories). (EX-1001

(’337 patent), claim 1; EX-1205 (Webster’s New World Dictionary), 120.) Regardless, in a device with a proximity sensor, a voice received from one distance is “separately identifiable” from a voice received from another distance.

9. Mr. McAlexander does not propose a construction for “classify” and I understand that the Patent Owner has not done so either. I do not believe a construction is necessary, since a person of ordinary skill in the art would have understood the term’s plain and ordinary meaning. However, were “classify” to be construed, it should be construed to mean “arrange in categories”, which is its plain and ordinary meaning. (EX-1205 (Webster’s New World Dictionary), 120.)

B. “Voice”

10. In Paragraph 44, Mr. McAlexander states that “[t]he word ‘voice’ does not equate to [a] mere incidental change in ‘volume’ due to a measurement device being moved away from a speaker.” (EX-2013 ¶44.) Further, in Paragraph 45, Mr. McAlexander refers to a “multivariate notion of voice.” (*Id.* ¶45.) While I agree that a voice may have many characteristics or variables, the issue here is which characteristic or variable may be used to classify the voice. The ’337 patent discloses that a voice may be classified based on a single variable, such as the volume or the speaker’s distance from the device. (EX-1001 (’337 patent), 7:64-8:34.)

11. My testimony in this proceeding is consistent with this understanding.

I have explained that a voice has many characteristics or variables, and that an increase in volume is usually associated with other changes that could alternatively be detected and used to classify a voice. (*E.g.*, EX-2015, 60:3-62:16.) This is entirely consistent with the '337 patent's disclosure.

12. I understand that the Patent Owner has argued that each of the variables discussed in the specification as variables used for classification are merely for purposes of determining whether a user's voice is a whisper or not a whisper. I disagree. While the specification discusses methods of determining whisper, its discussion of variables like volume and distance is not limited to whisper. (EX-1001 ('337 patent), 7:64-8:34.)

13. I understand the Patent Owner has also argued that under the Board's construction of the claims, every change in voice volume would be deemed to be a new voice (*e.g.*, a second voice). I disagree. The patent's specification is clear that the system may set "a threshold of the feature value" (*e.g.*, distance or volume), and may classify the voice "based on whether the feature value exceeds the threshold." (EX-1001 ('337 patent), 8:30-34.) Thus, for example, all volume changes below the threshold would not result in a different voice classification.

C. Analyze a Voice to Classify it Based on Distance

14. In Paragraph 56, Mr. McAlexander asserts that "the plain language of limitations 1[b], 4[b], and 5[b] expressly require **(1)** a classification that must be

based on (2) an analysis of the acquired voice and (3) a distance, which is calculated (e.g., by a proximity sensor) in limitations 1[a], 4[a], and 5[a].” (EX-2013 ¶56.) I disagree. The claims recite a device configured to “analyze a voice spoken by a user ... to classify the voice as either one of a first voice or a second voice based on the distance between the user” and the device. (EX-1001 (’337 patent), claim 1.)

15. I agree with Mr. McAlexander that the distance *calculation* is not the claimed analysis. The distance calculation is performed by the proximity sensor. The voice classifying unit then analyzes the voice by determining whether its distance exceeds a threshold. (EX-1001 (’337 patent), 8:20-34.) That analysis allows the classifying unit to “classify” the voice as being nearer than the threshold (a “first voice”) or farther than the threshold (“second voice”). (*Id.*) Thus, the analysis includes comparing the voice’s distance to the threshold value, and the result of that analysis is the classification of the received voice as a “first voice” or a “second voice.”

16. In Paragraph 57, Mr. McAlexander suggests that “[t]he specification consistently describes analyzing a user’s voice for purposes of determining whether the user’s voice is a whisper and classifying the user’s voice on that basis.” (EX-2013 ¶57.) I agree that the specification describes detecting whisper, but whisper is disclosed as merely one potential basis for classifying a voice and other embodiments classify the voice based on other variables, including distance. (EX-1001

(’337 patent), 7:51-8:34.) Indeed, a person of ordinary skill in the art would have understood the specification to disclose that voice can be classified “by using any method,” including distance alone. (*Id.*) And the claims do not require any classification based on whisper. (*Id.*, claims 1-5.) Thus, I disagree with Mr. McAlexander to the extent he suggests that the claims require classification based on something other than distance, such as whether the received voice is a whisper.

II. SHIN

A. Shin Discloses Classifying Voice Based on Distance.

17. In Paragraphs 68-74, Mr. McAlexander opines that Shin does not disclose voice classification. (EX-2013 ¶¶68-74.) I disagree. Shin expressly discloses “analyz[ing] a voice input[.]” (EX-1003 (Shin) ¶[0048].) Indeed, Shin discloses “receiv[ing] a speech of a user *as a voice input* ... to determine an output scheme ... from *an analysis of the voice input*[.]” (*Id.*, Abstract (emphases added).) Shin discloses analyzing a voice input—for example, evaluating speech rate, volume, presence of a keyword, and/or user-to-device distance—to alter the output, just as in the ’337 patent. (See, e.g., *id.* ¶¶[0049], [0053].) Shin’s Figures also plainly disclose analyzing a voice spoken by a user:

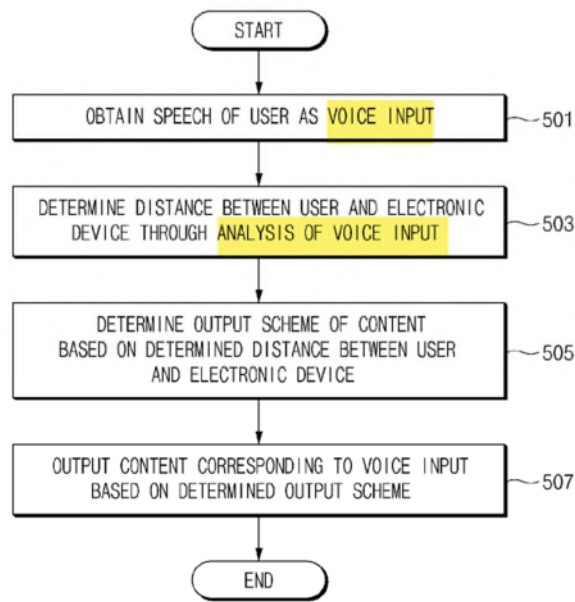


FIG. 5A

(*Id.*, Fig. 5A (annotated).)

18. In Paragraph 69, Mr. McAlexander opines that Shin’s “Table 3 discloses one thing: a calculation of distance—nothing else[.]” (EX-2013 ¶69.) I disagree. Table 3 does not disclose *calculation* of distance. Shin discloses that the distance is calculated by the proximity sensor and, optionally, the processor. (See, e.g., EX-1003 (Shin) ¶¶[0075]-[0077], claim 9.) Rather than calculations, Table 3 expressly discloses *classifications based on distance*.

TABLE 3

Distance (D) between user and electronic device	Output amount of information of content
$D < 1 \text{ m}$	Amount of information of original content
$1 \text{ m} \leq D < 2 \text{ m}$	Content of first abbreviation level
$2 \text{ m} \leq D < 3 \text{ m}$	Content of second abbreviation level
$3 \text{ m} \leq D < 4 \text{ m}$	Content of third abbreviation level
$D \geq 4 \text{ m}$	Content of fourth abbreviation level (abstract content)

(*Id.*, Table 3 (consolidated).) The left column of Table 3 includes several categories (or classifications) into which the distance D may fall. (*Id.*) And the right column of Table 3 describes the result of each such classification on the device’s output. (*Id.*)

19. This is exactly how the ’337 patent describes voice classification. The ’337 patent describes two classes: (1) $D \leq x$; and (2) $D > x$, where x is a threshold value. (EX-1001 (’337 patent), 8:30-34.) In the ’337 patent, the device receives a voice, calculates the user-to-device distance D , determines which distance-based category (class) the received voice belongs into, and the appropriate output is provided. (*Id.*, 8:60-66.) Shin discloses the same thing. The only difference is that the ’337 patent gives the two classes generic names: “V1A” (or “first voice”) and “V1B” (or “second voice”), whereas Shin does not expressly refer to the classes’ *names*. But Shin still discloses the classifications.

20. Similarly, I disagree with Mr. McAlexander’s assertion in Paragraph

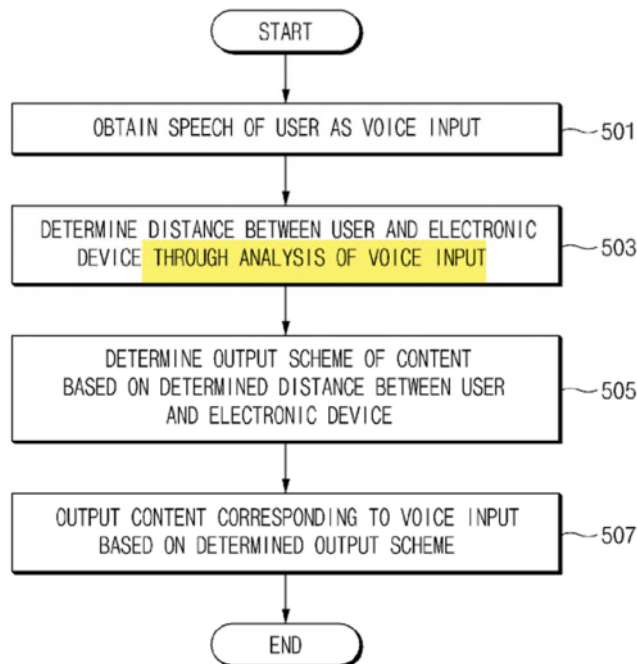
73 that “Shin directly correlates content output to distance” and thus “does not classify the voice input in any way[.]” (EX-2013 ¶73.) Shin’s device determines the distance between the user and the device, classifies the voice by comparing the user-to-device distance to the categories of distance ranges (e.g., the classes in Table 3), and outputs the content corresponding to the applicable class. Again, this is exactly what the ’337 patent describes and claims.

B. Shin Discloses Voice Analysis.

21. In Paragraphs 75-79, Mr. McAlexander opines that Shin does not disclose voice analysis. (EX-2013 ¶¶75-79.) I disagree. Shin plainly discloses analyzing the acquired voice to classify the voice based on distance. Shin explicitly discloses “analyzing” the voice input. (*See, e.g.*, EX-1003 (Shin) ¶¶[0037], [0048] (processor “analyze[s] a voice input”), [0076]-[0077] (user-to-device distance may be determined by “analyzing the voice”), Abstract.) Shin also explicitly discloses analyzing the received voice by comparing distance to the voice’s origination point to threshold values for the purpose of classifying the voice into one of several classes, so that the device can generate the appropriate output (e.g., content and volume). (*Id.* ¶¶[0076]-[0080].) This is the same “analysis” described in the ’337 patent, and the claims require no more.

22. I understand that the Patent Owner has argued that Shin does not determine user-to-device distance based on any voice analysis. I disagree. First, I note

that the challenged claims do not require the user-to-device distance determination to be based on an analysis of the voice; they require the user-to-device distance to be calculated by a proximity sensor. Second, even if the claims were to require determining distance based on an analysis of the voice’s sound signal, Shin discloses that. Shin discloses that the user-to-device distance can be calculated by a proximity sensor, “by analyzing the voice input” and determining the distance “based on the obtained volume of speech,” or by using both methods together to determine distance more accurately. (EX-1003 (Shin) ¶¶[0075]-[0077].) Again, Shin’s Figure 5A clearly discloses that the distance may be determined “through analysis of voice input”:



(*Id.*, Fig. 5A (annotated).)

23. I further understand that the Patent Owner has questioned whether the

processes disclosed in connection with Shin’s Tables 1 and 3 are compatible with one another. They are. Shin expressly discloses that the embodiments described in Tables 1 and 3 may be used together. (*E.g.*, EX-1003 (Shin) ¶¶[0049] (disclosing determining “the output scheme” which “may include an output volume level ... and an output amount of information”), [0053], [0140], claim 17 (disclosing “adjusting at least one of an output volume level of the content ... or an output amount of information of the content based on the distance between the user and the electronic device”); *see also id.* ¶¶[0183]-[0184].)

III. SHIMOMURA

A. Shimomura Discloses Voice Classification.

24. In Paragraph 88, Mr. McAlexander argues that Shimomura’s “Figure 7 discloses one and only one thing: a calculation of distance.” (EX-2013 ¶88.) I disagree. Shimomura’s Figure 7 discloses classification of voice input based on distance. I note that in Paragraph 87, Mr. McAlexander acknowledges that Figure 7 discloses “classes of distances.” (*Id.* ¶87.) Indeed, Figure 7 describes classes or categories of distances *from which the voice is received*. Thus, Shimomura discloses classifying a received voice based on whether the distance to the user exceeds a threshold (e.g., 350cm), exactly as the ’337 patent describes classifying a voice based on distance. (EX-1001 (’337 patent), 8:30-34 (“the voice classifying unit 38

... classifies the voice V1 ... based on whether the feature value exceeds the threshold”).) The top row of Figure 7 includes two discrete classifications into which the voice may fall. (EX-1004 (Shimomura), Fig. 7.) And the bottom row describes the result of each such classification on the device’s output. (*Id.*)

25. In Paragraph 91, Mr. McAlexander concludes that Shimomura “directly correlates the output content and volume based on distance without performing the step of classifying the voice.” (EX-2013 ¶91.) I disagree. Shimomura uses the distance to classify the received voice as coming from less than a threshold distance or more than a threshold distance. The result is used to determine the content and volume of the output. Again, this is precisely what the ’337 patent describes and claims. (EX-1001 (’337 patent), 8:30-34.)

B. Shimomura Discloses Voice Analysis.

26. In Paragraph 95, Mr. McAlexander opines that “Shimomura does not disclose determining user-to-device distance based on any voice analysis.” (EX-2013 ¶95.) I disagree that this observation is relevant. Nothing in the ’337 patent’s claims requires the distance determination to be based on a voice analysis. Indeed, the ’337 patent never describes determining distance based on any analysis of the voice signal; it describes determining distance only with a proximity sensor. (EX-1001 (’337 patent), 8:20-34.) Thus, Shimomura describes calculating the user-to-device distance in the same way as the ’337 patent, and also describes classifying

the voice based on that distance in the same way, i.e., whether the distance is greater or less than a threshold value. (*Id.*; EX-1004 (Shimomura) ¶[0093].)

C. Shimomura Renders the Claims Obvious.

27. In Paragraph 98, Mr. McAlexander asserts that “Shimomura does not expressly disclose a connection between the processes associated with Figures 7 and 9.” (EX-2013 ¶98.) I disagree. As part of the process described in Shimomura, the device receives a “character string” of potential output, and based on the distance, classifies the input, determines whether to alter the output character string, and then passes the resulting (potentially altered) character string to be output. (EX-1004 (Shimomura) ¶¶[0059]-[0066], Figs. 7-8.) Then, Shimomura’s device again uses the distance-based classification to determine whether to alter the volume of the output. (*Id.* ¶¶[0066]-[0070], Figs. 9, 13.) Shimomura explicitly discloses that the first process is the input to the second; thus, they are used in conjunction.

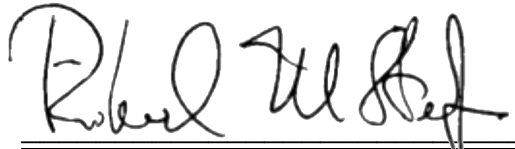
28. In Paragraph 107, Mr. McAlexander alleges that Shimomura’s omission of “ne” is “a ‘change[] in intonation’” and “not an omission of information.” (EX-2013 ¶107.) I disagree. Shimomura does not specify that it is talking about “ne” when it discusses change in intonation. (EX-1004 (Shimomura) ¶[0092].) And Shimomura also discloses that “words other than ‘ne’ can be used as words to be inserted,” so Shimomura’s disclosure is not limited as Mr. McAlexander argues. (*Id.*)

29. In Paragraph 115, Mr. McAlexander argues that “the authors of Shin and Shimomura specifically declined to conduct an analysis of voices to classify voices, let alone based on distance” and “purposely chose simpler, more discrete approaches.” (EX-2013 ¶115.) I disagree. Shin and Shimomura disclose classifying voice input by distance exactly the same way as the ’337 patent does, as I describe above and in my prior declaration. No added complexity is required to combine or modify those references because the systems already work the same way; in fact, their simplicity supports their combination.

Amazon.com, Inc. v. SoundClear Technologies LLC
Reply Declaration of Richard Stern, Ph.D. – IPR2025-00565

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

Executed on March 2, 2026 at Buenos Aires, Argentina.



Richard Stern, Ph.D.