UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

APPLE INC.,

Petitioner

v.

PARUS HOLDINGS, INC.,

Patent Owner

IPR2020-00686

Patent No. 7,076,431

AND

IPR2020-00687

Patent No. 9,451,084

SUPPLEMENTAL DECLARATION OF DR. LOREN TERVEEN

I, Dr. Loren Terveen, hereby declare the following:

I. INTRODUCTION

1. I have been asked to respond to certain issues raised by Patent Owner in Patent Owner's Response dated December 23, 2020 ("POR"). All of my opinions expressed in my original declaration (Ex. 1003) remain the same. I have reviewed the relevant portions of the POR (Paper 15) and the relevant portions of Mr. Occhiogrosso's declaration (Ex. 2025) and deposition transcript (Ex. 1039) in connection with preparing this supplemental declaration. References to opinions of the '431 Patent below are intended as equally applicable to the '084 Patent.

II. OPINIONS

A. A Two-Step Speech Recognition Process Is Described in Both the '431 and '084 Patents and Ladd

2. As I discussed in my original declaration (Ex. 1003) at ¶¶ 81-83, *Ladd* teaches a system for retrieving information by uttering speech commands into a voice enabled device and for providing information retrieved from an information source, such as "web pages" or "web sites." Specifically, *Ladd's* system is an IVR (Interactive Voice Response) system that may answer a question, such as "what is the weather" from a web site in response to a spoken user request. Ex. 1003, ¶ 78, 81-82 *citing Ladd*, 2:19-64, 3:7-53, 9:1-21. In an IVR system, including specifically *Ladd's*, the computing system must determine the content of at least some of the speech uttered by the user in order to identify desired information for retrieval from

an appropriate information source. For example, when a user inquires about the current weather in Chicago, the system must determine the key words "weather" and "Chicago" were spoken and by comparison to the grammar, determine the command corresponding to the spoken words, i.e., that the user is commanding to retrieve Chicago's weather. *Ladd*, 2:48-54, 4:64-5:11, 8:23-25, 10:3-11, 11:50-64, 38:4-16. This is in contrast to Mr. Occhiogrosso's description of mere transcribing of free speech that may occur in some systems, where spoken utterances are transformed from audio messages into text and stored in memory, but no content is determined for any transcribed words. Ex. 1039, *Occhiogrosso Dep. Tr.,* 39:10-40:22.

3. In order for an IVR system to act upon user speech, it must perform two distinct steps. In the first step, the speech recognition device simply transforms the sound wave into text. Ex. 1039, 33:11-16, 49:5-19. At this juncture, the speech recognition device has not yet determined any content of what was said, i.e., what instruction is being commanded; it has merely generated a textual data message. *Id.* For example, a speech recognition device that has performed only this first step may generate the character string "weather" after the word "weather" was spoken, but the device does not yet know what to do in response to the character string "weather." There are a number of methods by which a system may perform this first step of converting the spoken words into text, but *Ladd* is not specific on how it requires step one to occur. I note that Mr. Occhiogrosso also agrees there are various speech

recognition algorithms to recognize the user's speech and convert into text. Ex. 1039, 54:6-16.

4. It is not until the second step of content recognition of the spoken speech that a speech recognition device determines the content of the spoken words (e.g., determining that the user uttered "weather" and is therefore instructing the IVR system to retrieve and respond with the current weather). Mr. Occhiogrosso agreed with this during his deposition in differentiating between the first step of converting speech into text and the second step of using a recognition grammar to "address[] what words are." *Id.* at 50:17–51:8. Speech recognition devices that do not determine the content of transcribed words cannot act in response to the spoken words. *Id.* at 40:13-22. (Mr. Occhiogrosso opining that when the user is "simply speaking and there is no higher order context of a recognition grammar that meters or governs the speech, then the speech recognition engine will dutifully translate what the user is speaking into text" and that "free speech" or "free text" is "effectively a dictation application with no imposed recognition grammar"). Systems such as the '431 Patent and *Ladd* must perform **both** steps to act upon a spoken command to retrieve desired information, namely the steps of (1) converting speech utterances into text words, and (2) comparing the textual words to grammar to determine the content of the spoken command. As I explain further below, the statement in the '431 Patent that it "recognizes spoken words *without* using predefined voice patterns" is

characterizing a method of performing the **first** step of speech recognition (transforming speech to text). In contrast, *Ladd's* description of determining a "speech pattern" is characterizing a method of performing the **second** step of speech recognition (determining the *content* of the text). *'431 Patent*, 4:38-43; *Ladd*, 9:27-44. I further note this second step is recited in the claims of the '431 Patent at Limitations 1(f)-1(h), which recite the recognition grammar, that the speech command comprises an information request selectable by the user, and selecting the recognition grammar upon receiving the speech command.

5. The '431 Patent confirms the two-step process. Specifically, the speech recognition engine 300 "converts voice commands received from the user's voice enabled device 112…into data messages." *'431 Patent*, 6:4-8. "The media server 106 uses the speech recognition engine 300 to interpret the speech commands received from the user. Based upon these commands, the media server 106 retrieves the appropriate web site record 200 from the database 100." *Id.* at 16:3-7. Therefore, the '431 Patent describes a system where the speech commands are converted into data messages, i.e., text, and then the converted speech commands are interpreted to determine what web site record to retrieve.

6. Ladd also confirms its system performs a two-step speech recognition process, stating: "The STT unit 256 of the VRU server 234 receives speech inputs or communications from the user and converts the speech inputs to textual

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.