**UNITED STATES PATENT AND TRADEMARK OFFICE**

---

**BEFORE THE PATENT TRIAL AND APPEAL BOARD**

---

**SAMSUNG ELECTRONICS CO. LTD. and SAMSUNG ELECTRONICS AMERICA, INC.**
***Petitioners***

**v.**

**VB ASSETS, LLC,**
***Patent Owner***

---

**IPR2025-00867**
**U.S. Patent No. 7,818,176**

---

# DECLARATION OF DR. ANATOLE GERSHMAN IN SUPPORT OF PATENT OWNER'S PRELIMINARY RESPONSE

**TABLE OF CONTENTS**

Page

I. INTRODUCTION AND ENGAGEMENT ..... 1

II. QUALIFICATIONS ..... 3

III. LEGAL PRINCIPLES ..... 6

IV. SUMMARY OF OPINIONS ..... 10

V. LEVEL OF ORDINARY SKILL IN THE ART ..... 11

VI. PROSECUTION HISTORY OF THE '176 Patent ..... 12

VII. REBUTTAL TO GROUND 1 ..... 17

A. Overview of Asserted Art ..... 17

1. Taschereau ..... 17

2. Hansen ..... 20

B. Petitioners Fail to Demonstrate Motivation to Combine Taschereau with Hansen ..... 23

C. Claims [1B]/[27B] ..... 31

D. Claims [1D]-[1E]/[27D]-[27E] ..... 33

VIII. REBUTTAL TO GROUND 2 ..... 36

A. Overview of Asserted Art ..... 36

1. Taschereau and Hansen ..... 36

2. Franco ..... 37

B. Petitioners Fail to Demonstrate Motivation to Combine Taschereau-Hansen with Franco ..... 38

C. Independent Claims 14 and 40 and Claims 12, 38 ..... 41

IX. REBUTTAL TO GROUND 3 ..... 42

X. REBUTTAL TO GROUND 4 ..... 43

A. Overview of Asserted Art ..... 43

1. Ramer ..... 43

2. Cross ..... 46

B. Independent Claims [1C]/[27C]. ..... 46

C. Independent Claims 14 and 40. ..... 51

XI. CONCLUSION ..... 56

I, Anatole Gershman, do hereby declare and say as follows:

## I. INTRODUCTION AND ENGAGEMENT

1. I have personal knowledge of the facts contained in this Declaration, am of legal age, and am otherwise competent to testify.

2. I have been retained by counsel for VB Assets LLC ("Patent Owner"), as an independent technical expert in the above-captioned proceeding. I have been asked to provide my opinion about the state of the art of the technology described in U.S. Patent No. 7,818,176 ("the '176 patent") and on the patentability of the claims of this patent.

3. The following is my written testimony on these topics. The opinions contained in this Declaration are based on the documents I reviewed, as well as my education, years of professional experience, and knowledge. The materials I have considered in reaching these opinions include the '176 patent, the petition (Paper No. 3, hereinafter the "Petition") and the exhibits or other material cited therein (*i.e.*, Exhibits 1001-1032), as well as Patent Owner's Discretionary Denial Brief (Paper No. 8), the exhibits or other material cited therein, and any additional exhibits or other material cited in this Declaration.

4. This Declaration presents the opinions I have formed at the present time, and I reserve the right to revise, supplement, or amend my opinions stated in this Declaration based on any new information that I may be asked to review.

5. I understand that Samsung Electronics Co. Ltd. and Samsung Electronics America, Inc. (collectively, "Petitioners" or "Samsung") filed a petition for *Inter Partes* Review ("IPR") of the '176 patent, in which Petitioners assert that:

- Claims 1-2, 9-11, 27-28, and 35-37 are obvious over International Publication WO2004/055781 to Taschereau et al. ("Taschereau"; EX-1005) in view of U.S. Patent 5,640,490 to Hansen et al. ("Hansen"; EX-1006) (Ground 1);
- Claims 12-18, 25-26, 38-44, and 51-52 are obvious over Taschereau and Hansen, and further in view of U.S. Publication 2010/0125458 to Franco et al. ("Franco" or "EX-1007") (Ground 2);
- Claims 3-8, 19-24, 29-34, and 45-50 are obvious over Taschereau and Hansen, and further in view of U.S. Publication 2007/0174258 to Jones et al. ("Jones"; EX-1008) (Ground 3); and,
- Claims 1-52 are obvious over U.S. Patent 7,577,665 to Ramer et. al. ("Ramer"; EX-1009) in view of U.S. Patent 8,332,218 to Cross et. al. ("Cross"; EX-1010) (Ground 4).

6. I have been asked, for purposes of Patent Owner's Preliminary Response, to analyze and respond to particular assertions in the Petition. I reserve the right to analyze and respond to other alleged patentability challenges if this proceeding is instituted.

7. As explained below, in my opinion, the Petition fails to show that claims 1-52 of the '176 Patent are rendered obvious.

8. I am being compensated at a rate of $500 per hour for my study and other work in this matter. I am also being reimbursed for reasonable and customary expenses associated with my work in this investigation. My compensation is not contingent on the outcome of this matter or the specifics of my testimony.

## II. QUALIFICATIONS

9. My curriculum vitae, attached here as EX-2026, lists my qualifications. To summarize, I recently retired from my position as a Distinguished Career Professor in the Language Technologies Institute at Carnegie Mellon University, a position which I had held for 16 years. After retiring from CMU, from July 2023 through June 2024 I consulted with the University of Colorado, Department of Linguistics, regarding a DARPA project that has since concluded. I received my Bachelor of Science and Master of Science Degrees in Mathematics and Computer Science from Moscow State Pedagogical University in 1971. I received a Ph.D. in

Computer Science from Yale University in May 1979. My dissertation was entitled "Knowledge-based Parsing" and focused on natural language processing.

10. My first job after graduation was as a researcher (Member of Technical Staff) at Bell Laboratories focusing on natural language access to Yellow Pages. In the 1980s I was a Vice-President of Cognitive Systems, Inc, a company dedicated to commercialization of natural language processing. At Cognitive Systems, Inc., we built systems that extracted information from natural language money transfer telexes and other financial documents for major banks, including Citibank and Union Bank of Switzerland. These systems were put in operation. We also built prototypes of interactive natural language kiosks that sold financial products such as insurance and investments as well as helped customers manage their investment portfolios. These prototypes were created for Citibank and Societe Generale de Banque, a major Belgian Bank.

11. From 1989 through 2006 I was leading applied research projects at Accenture, first as Director of Human Systems Integration Laboratory and later as Chief Scientist and Global Director of Research. At Accenture, we focused on the emerging Internet and Ubiquitous Computing applications creating prototype demonstrations of how these technologies could create revolutionary changes in a range of industries. In 1994, we launched the first internet shopping agent that helped users find the best deal on music albums. Our prototypes included interactive natural

language home appliances such as medicine cabinets, wardrobes, health mirrors, smart kitchens, etc., which were featured in national and international news media. These systems featured automated interactive spoken dialog capabilities combined with a variety of physical sensors that helped users maintain their health, pick the right clothing for the occasion, create healthy menus for the family, etc. However, due to the limitations of the speech recognition technologies at that time, practical applications were deemed premature.

12. In 2007, after my retirement from Accenture, I was invited to join the faculty of Language Technologies Institute, a department at the School of Computer Science at Carnegie Mellon University in Pittsburgh, PA as Distinguished Career Professor, the position I held until my retirement in June of 2023. At Carnegie Mellon University, I participated in a number of research projects sponsored by major US Government agencies such as DARPA, IARPA and NMEC as well as companies such as PriceWaterhouseCoopers and Accenture. In addition, I consulted with several major companies such as Bank of New York/Mellon and MetLife. All of these projects involve extensive use of natural language processing. In-depth extraction and interpretation of information from natural language documents is the central challenge of the two DARPA projects in which I participated. Specifically relevant to spoken dialog systems were the projects involving the analysis of customer service phone calls at MetLife, speech interactions with robotic vehicles

sponsored by the National Robotics Engineering Center – NREC at CMU, and a dozen of student projects exploring "in-depth" conversations with artificial agents.

13. After retiring from CMU in June of 2023, I consulted with the University of Colorado, Department of Linguistics on a DARPA project through June 2024.

14. During my 45 years of professional experience since earning my Ph.D., my focus has always been on intelligent systems and, in particular, on the knowledge that is necessary for systems to interact intelligently with their users. I have written more than 70 publications in the areas of natural language processing, the human-to-machine interface, artificial intelligence, and data science. I hold 13 US patents related to the human-to-machine interface, agent-based systems, customized media services, multimedia agents, e-commerce, and mobile object tracking.

## III. LEGAL PRINCIPLES

15. For my opinions in this declaration, I understand that it requires applying various legal principles. As I am not an attorney, I have been informed and understand from counsel about various legal principles that involve my analysis. I have used my understanding of those principles in forming my opinions. I summarize those principles as I understand them below.

16. For example, I understand that Petitioners bear the burden of proving unpatentability in this proceeding by a preponderance of the evidence. I understand

that this preponderance of the evidence standard means that Petitioners must show that unpatentability is more probable than not

17. I understand that a claimed invention is not patentable under 35 U.S.C. § 103 if it is obvious. A patent claim is unpatentable if the claimed invention would have been obvious to a person of ordinary skill in the field (which I discuss below) at the time the claimed invention was made. This means that even if all of the requirements of the claim cannot be found in a single prior art reference that would anticipate the claim, a person of ordinary skill in the relevant field who knew about all this prior art would have come up with the claimed invention.

18. I have been informed and understand that the obviousness analysis requires evaluating the claim language as a whole, including comparing each and every limitation of the properly construed claim language to the prior art.

19. I further understand that the ultimate conclusion of whether a claim is obvious should be based upon several factual determinations. That is, a determination of obviousness requires inquiries into: (1) the level of ordinary skill in the field; (2) the scope and content of the prior art; (3) what difference, if any, existed between the claimed invention and the prior art; and (4) any objective factors bearing on obviousness.

20. I understand that, in determining the level of ordinary skill in the field that someone would have had at the time the claimed invention was made, I should

consider: (1) the levels of education and experience of persons working in the field; (2) the types of problems encountered in the field; and (3) the sophistication of the technology.

21. I understand that, in determining the scope and content of the prior art, in order to be considered as prior art, a reference must be analogous art. I also understand that to be analogous art, the reference must be either (1) from the same field of endeavor as the claimed subject matter, regardless of the problem addressed, or (2) if not in the same field of endeavor, reasonably pertinent to the particular problem with which the inventor is involved. I am also familiar with the premise that for a reference to be reasonably pertinent, it must have logically commended itself to an inventor's attention at the time of invention.

22. I understand that a patent claim composed of several elements is not proved obvious merely by demonstrating that each of its elements was independently known in the prior art. In evaluating whether such a claim would have been obvious, I may consider whether there is a reason that would have prompted a person of ordinary skill in the field to combine the elements or concepts from the prior art in the same way as in the claimed invention.

23. I understand that in order to show obviousness based on a single reference or a combination of references, a particular motivation to modify the

reference or combine the teachings in the references, and a reasonable expectation of success must be shown.

24. I understand that the art must evidence a motivation to combine or modify the independently known elements to arrive at the claimed invention and an explanation as to how or why the references would be combined to arrive at the claimed invention to solve the particular problem. I understand that a challenger must show that there was a reason that would have prompted a person of ordinary skill in the relevant field to combine the elements in the way the claimed new invention does in an obviousness determination. I also understand that it is not enough to prove obviousness merely by a showing that the references were capable of combination without a suggestion of the desirability of the modification.

25. I further understand that there is no single way to define the line between true inventiveness on the one hand (which is patentable) and the application of common sense and ordinary skill to solve a problem on the other hand (which is not patentable). For example, market forces or other design incentives may be what produced a change, rather than true inventiveness. I may consider whether the change was merely the predictable result of using prior art elements according to their known functions, or whether it was the result of true inventiveness. I may also consider whether there is some teaching or suggestion in the prior art to make the modification or combination of elements claimed in the patent. I may consider

whether the innovation applies a known technique that had been used to improve a similar device or method in a similar way. I may also consider whether the claimed invention would have been obvious to try, meaning that the claimed innovation was one of a relatively small number of possible approaches to the problem with a reasonable expectation of success by those skilled in the art.

26. I also understand, however, that I must be careful not to determine obviousness using the benefit of hindsight; many true inventions might seem obvious after the fact. I should put myself in the position of a person of ordinary skill in the field at the time the claimed invention was made and I should not consider what is known today or what is learned from the teaching of the patent.

27. Finally, I understand that any obviousness rationale for modifying or combining prior art must include a showing that a person of ordinary skill would have had a reasonable expectation of success. I further understand that whether a proposed modification or combination of the prior art has a reasonable expectation of success is determined at the time the invention was made.

## IV. SUMMARY OF OPINIONS

28. In my opinion, Petitioners' proposed combination of Taschereau and Hansen does not render obvious claims 1-2, 9-11, 27-28, and 35-37 of the '176 patent (Ground 1).

29. In my opinion, Petitioners' proposed combination of Taschereau, Hansen, and Franco does not render obvious claims 12-18, 25-26, 38-44, and 51-52 of the '176 patent (Ground 2).

30. In my opinion, Petitioners' proposed combination of Taschereau, Hansen, and Jones does not render obvious claims 3-8, 19-24, 29-34, and 45-50 of the '176 patent (Ground 3). Ground 3 relies on the combination of Ground 1, further modified by Jones, but does not challenge any independent claims. Petitioners do not allege that Jones resolves or addresses any of the functionality relied upon in Ground 1 for the independent claims challenged therein. Therefore, my opinions concerning Petitioners' failure to demonstrate Ground 1 renders any of the challenged independent claims obvious applies equally to all claims challenged in Ground 3.

31. In my opinion, Petitioners' proposed combination of Ramer and Cross does not render obvious claims 1-52 of the '176 patent (Ground 4).

## V. LEVEL OF ORDINARY SKILL IN THE ART

32. I understand that Petitioners provide the following definition of a POSITA:

> A person of ordinary skill in the art ("POSITA") would have had a bachelor's degree in computer science, computing engineering, electrical engineering, or related field in computing technology, and two years of experience with automatic speech recognition and natural

> language understanding, or equivalent education, research experience, or knowledge.

Pet. at 6.

33. For purposes of this Declaration in support of Patent Owner's Preliminary Response, I offer no independent opinion of the level of ordinary skill in the art but have adopted the definition proposed by Petitioners. I reserve the right to further address the level of ordinary skill in the art if this proceeding is instituted.

34. Each of the arguments below should be considered from the perspective of a POSITA in the field of the '176 Patent and during the relevant time frame (e.g., February 6, 2007). During that time frame, as described above and as illustrated in my CV (*see* EX-2026), I had at least the requisite level of skill and experience to be a person of ordinary skill in the art. Furthermore, I am familiar with the level of skill in the art that a POSITA would have had at the time, as during the relevant time frame I regularly supervised, interacted with, and taught persons who had this level of skill.

## VI. PROSECUTION HISTORY OF THE '176 PATENT

35. I have reviewed the prosecution history of the '176 Patent (EX-1002). Application No. 11/671,526, which later issued as the '176 Patent, was filed on February 6, 2007. *See* EX-1002, 1-34.

36. On November 20, 2009, the Office issued a Non-Final Office Action. EX-1002, 149. In the Office Action, the then pending claims were rejected under 35 U.S.C. § 101 for being "directed to non-statutory subject matter" and 35 U.S.C. § 102(e) as being anticipated by Richardson et al. (U.S. Pat. No. 7,461,059). EX-1002, 151-158.

37. In response, on April 20, 2007, the Applicant amended the claims, added new claims 21-28, disputed that the claims were "directed to non-statutory subject matter," and distinguished the Examiner's arguments over Richardson. EX-1002, 181-193. Specifically, the Applicant disputed that the Richardson reference disclosed "at least the combination of features that includes 'recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device,' 'interpreting the recognized words or phrases . . . [to. establish] a context for the natural language utterance,' and 'selecting an advertisement in the context established for the natural language utterance.'" EX-1002, 191.

38. In response the Office Issued a Final Office Action on May 27, 2010. EX-1002, 203-227. In the Final Office Action, the Examiner agreed that the "Applicant's arguments are persuasive" to overcome the "rejections made under 35 U.S.C. 101 in the previous Office Action" but objected to Claims 3, 11, 13, and 23 (for certain alleged informalities), rejected Claims 4, 5, 7, 10, 14, 15, 17, 20, 24, 25,

27, and 28 under 35 U.S.C. § 112, first paragraph, as failing to comply with the written description requirement, rejected Claims 1-3 and 11-13 under 35 U.S.C. § 102(e) as being anticipated by Atsom et al. (U.S. Pat. Publ. No. 2010/0063880), rejected Claims 4-6, 8, 9, 14-16, 18, and 19 under 35 U.S.C. § 103(a) as being unpatentable over Atsom, in view of Maislos et al. (U.S. Pat. Publ. No. 2007/0186165), and rejected Claims 10, 20, 21, and 25 under 35 U.S.C. § 103(a) as being unpatentable over Atsom in view of Kennewick et al. (U.S. Pat. Publ. No. 2004/0193420). EX-1002, 205-224. Claims 22-24 and 26-28 were objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form. EX-1002, 224-225. In stating that Claims 22-24 and 26-28 were allowable, the Final Office Action stated: (1) "[w]ith respect to claim 22, the prior art of record does not disclose mapping a stream of phonemes to one or more syllables that are phonemically represented in an acoustic grammar and generating a preliminary interpretation from the one or more syllables (see related U.S. Patent 7,634,409 assigned to the instant application's assignee);" (2) [w]ith respect to claims 23 and 24, the prior art of record does not disclose or suggest determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event (see related U.S. Patent 7,620,549

assigned to the instant application's assignee);" (3) "Claim 26 is similar in scope to claim 22 and is allowable for the same reasons as claim 22;" and (4) "Claims 27 and 28 are similar in scope to claims 23 and 24 and are allowable for the same reasons as claims 27 and 28." EX-1002, 224-225.

39. In response, on July 22, 2010, the Applicant filed a Reply and Amendment Accompanying Request for Continued Examination in which the Applicant amended the specification, amended Claims 2-5, 7, 10, 12-15, 17, 20-23, and 25-27, cancelled Claims 1 and 11, added Claims 29-54, and traversed the rejections. EX-1002, 238-260. Specifically, the Applicant rewrote Claims 22, 23, 26, and 27 in independent form so that these claims "recite[d] subject matter that the Examiner has identified as allowable." EX-1002, 256. The Applicant then explained that "[C]laims 2-10, 12-21, 22, 24-25, and 28-54 depend from and add features to one of amended independent claims 22, 23, 26, and 27, whereby claims 2-10, 12-21, 22, 24-25, and 28-54 therefore recite allowable subject matter for at least the reason that they depend from independent claims 22, 23, 26, and 27, which recite subject matter that the Examiner has identified as allowable." *Id*. The Applicant also amended Claims 3, 13, and 23 to address the alleged typographical errors that the Examiner identified. The Applicant also argued that the amendments to the specification provide written description for the features recited features in Claims 4-5, 7, 10, 14-15, 17, 20, 24-25, and 27-28.

40. In the Reply and Amendment Accompanying Request for Continued Examination, the Applicant also traversed the rejection of Claims 1-3 and 11-13 under 35 U.S.C. § 102(e) because "previously pending claims 1 and 11 have been cancelled, whereby the rejection has been rendered moot with respect to claims 1 and 11" and because "the rejection with respect to claims 2-3 and 12-13 is improper and must be withdrawn for at least the reason that claims 2-3 and 12-13 recite subject matter that the Examiner has identified as allowable." EX-1002, 258. Similarly, the Applicant argued that the 35 U.S.C. § 103 "rejections [asserting the combinations of Atsom in view of Maislos and Atsom in view of Kennewick] are improper and must be withdrawn for at least the reason that claims 4-6, 8-10, 14-16, 18-21, and 25 recite subject matter that the Examiner has identified as allowable." EX-1002, 259.

41. The Office subsequently issued a Notice of Allowance on August 11, 2010. *See* EX-1002, 270-275. The Examiner explained that "[t]he amendments made to the specification overcome the rejections under 35 U.S.C. 112, 1st paragraph made in the previous Office Action," "new claims 29-54 find support in the specification, either directly, or through an incorporated by reference application in accordance with MPEP 2163.07(b)," and "[w]ith respect to the prior art, claims 22, 23, 26, and 27 have been rewritten in independent form including all limitations of the base claim and any intervening claims. Accordingly, claims 22, 23, 26, and 27 are now allowable. The remaining claims depend from one of claims 22, 23, 26, and

27 and are therefore also allowable." EX-1002, 274. After an additional filing, a subsequent Notice of Allowance was issued and the patent issued on October 19, 2010. *See* EX-1002, 303-305, 306-309, 314.

## VII. REBUTTAL TO GROUND 1

42. For the reasons below, it is my opinion the Taschereau-Hansen combination does not render obvious claims 1-2, 9-11, 27-28, and 35-37.

### A. Overview of Asserted Art

#### 1. Taschereau

43. Taschereau discloses voice recognition used in the context of directory assistance. EX-1005, 1 ("Technical Field").

44. Taschereau explains at "at this time and in the example provided, most ASR systems used in directory assistance applications are used exclusively with business listings, although they [sic] ASR systems can also be used with government of residential listings." EX-1005, 17:9-12.

45. Taschereau sets out to solve the "limitations of ASR systems" used "[i]n common practice" where "ASR is applied as a 'one shot' process whereby the ASR system is applied 'live' while the person is speaking and expected to return a result within a 'reasonable' period of time." EX-1005, 3:19-21; *see also id*., 6:12-14 ("In common practice, a grammar is generated and designed for 'single execution'.

That is, a grammar is generated knowing that the ASR technology will perform a 'single pass' on the grammar.").

46. Taschereau details many limitations with prior art ASR systems and directory assistance systems, including reduced accuracy with large grammars, timeouts in large grammars, loading issues, need for grammar splitting, pre-compiled grammar limitations, slower matching time, word order sensitivity, unnatural dialog flow, background noise disruptions, and inconsistent user experience. *See, e.g.*, EX-1005, 2:1-4:18.

47. Taschereau notes that "[t]here have been attempts to overcome these limitations of ASR systems by improving the efficiency of a single pass." EX-1005, 4:20-21.

48. To solve these problems and "address functional and accuracy problems associated with using ASR systems in general, and in particular, cases where large ASR 'grammars' are required," ***rather than improve the efficiency of a single pass like the prior art*** or use large grammars, "the process according to the [Taschereau] invention is a 'multi-pass' process." EX-1005, 5:1-2, 13:8-10 (emphasis added).

49. Accordingly, Taschereau's "grammar is designed to be as small as possible" and "[t]he grammar is dynamically generated knowing that the ASR system will be used again to perform one or more latent, and optionally concurrent,

recognitions, each latent recognition evaluating the terms from a previous recognition process." EX-1005, 6:16-19.

50. Taschereau explains that "when using ASR, the stored speech is re-applied in a manner which the caller cannot hear." EX-1005, 13:16-17. Likewise, in Taschereau's preferred embodiment, "the utterance is not re-presented to the user (i.e.: the user does not hear the original utterance even though it is used more than once)." EX-1005, 7:10-11.

51. A POSITA would have understood that Taschereau achieves its "multi pass" process by literally playing audio recordings in data form to the recognizer. Indeed, Taschereau explains that "[t]he use of a recorded utterance also allows multiple passes to be applied quickly through the grammars" because "the utterance will be in data form [so] it can be very quickly input into the recognizer (i.e. a five second utterance will not take twenty seconds to pass through four times)." EX-1005, 7:20-22.

52. Taschereau therefore discloses "matching an utterance" with "a listing in a directory" using a multi-pass process that "requires" the utterance to be "re-applied in a manner which the caller cannot hear." EX-1005, Abstract, 13:8-19.

53. Taschereau consistently references an "operator," which is a human, that is a core component of its disclosed directory assistance system to resolve situations where "the ASR system is not suitable for a response." EX-1005, 17:3-7.

In such a scenario, "the questions are whispered to the operator" or "[i]f the confidence of the ASR generated answer is not high enough . . ., then the responses to the questions can be whispered to an operator." EX-1005, 17:6-16.

54. For instance, Taschereau also discloses that "[t]he utterance, amended or unaltered, may also be 'whispered' to an operator at [the acoustic analysis and rendering] stage if the utterance has certain qualities that make it unsuitable for the ASR system, for example a large amount of background noise." EX-1005, 20:8-10.

**2. Hansen**

55. Hansen discloses that "the system and method of the current invention act as a means for entering data and/or commands to a standard personal computer" so as to "essentially replace[], or supplement[] other computer input devices, such as keyboards and pointing devices." EX-1006, 28:21-30.

56. Hansen does so through speech recognition and discloses a system that "includes a microphone and associated conditioning circuitry, for receiving an audio speech signal and converting it to a representative electrical signal" which is then "converted to a digital audio signal with a digital-to-analog converter." EX-1006, Abstract. Hansen evaluates characteristics of the signal and "[these characteristics are input to a programmable host sound processor which compares the sound characteristics to standard sound data." *Id*. Hansen "identifies the specific phoneme sounds that are contained within the audio speech signal" based on this comparison.

*Id*. Hansen's system also "includes linguistic processing program methods to convert the phoneme sounds into English words or other natural language words" and "[t]hese words are input to a host processor, which then utilizes the words as either data or commands." EX-1006, Abstract.

57. Hansen's "primary object" is a system "for achieving speech recognition." 2:11-13.

58. Hansen carries out this "primary object" with a "host sound processor" that includes a programmable device that is "programmed to carry out the steps necessary to evaluate the various sound characteristics contained within the data structure, and then generate the phoneme sound type that corresponds to those particular characteristics," then "performs the program steps needed to implement the linguistic processing portion of the overall method," and finally Hansen's "electronic means" receives this "word or phrase translated from the incoming stream of identified phonemes and "programmably processes the word as either data input, as for instance text to a wordprocessing application, or as a command input, as for instance an operating system command." EX-1006, 3:35-42.

59. Hansen discloses that the "host sound processor" functionality described above is carried out by its "Linguistic Processor." EX1-006, 9:66-10:11. Hansen explains that "[t]he Linguistic processor receives the data structures, and translates the sound stream (i.e., stream of phonemes) into the corresponding English

letter, syllable, word or phrase . . . by performing a variety of linguistic processing functions that match the phonemic sequences against entries in the system lexicon . . . include[ing] a phonetic dictionary look-up, a context checking function and database, and a basic grammar checking function." *Id*.

60. Hansen discloses that the "host sound processor" functionality described above is carried out by its "Command Processor." EX-1006, 10:12-19. Hansen "immediately" passes words that its Linguistic Processor extracts from the incoming speech signal to the Command Processor. EX-1006, 27:34-38.

61. Hansen's Command Processor takes each "particular word or phrase" and "determines whether the word or phrase constitutes text that should be passed as data to a higher level application, such as a wordprocessor, or whether it constitutes a command that is to be passed directly to the operating system or application command interface." EX-1006, 10:12-19, Fig. 11.

62. Hansen's Command Processor "determines, at step 452, whether the word received is an operating system command . . . by comparing the word to the contents of a definition file database, which defines all words that constitute operating system commands." EX-1006, 27:55-58. If the Command Processor determines that a command word is received, Hansen "passe[s] [the word] directly to the host computer 22 operating system" but if not, "it is determined if the word is instead an application command, as for instance a command to a wordprocessor or

spreadsheet . . . by comparing the word to another definition file database, which defines all words that constitute an application command" and "[i]f the word is an application command word, then it is passed directly, at step 458, to the intended application." EX-1006, 27:29-28:3.

63. Hansen explains that if "the Command Processor is still in a 'command mode'" and "the incoming word is neither a operating system command, or an application command" then "the word is discarded at step 464, and essentially ignored." EX-1006, 28:4-11. "If the Command Processor is not in a command mode, then the word will be sent directly to the current application as text." EX-1006, 28:4-11.

64. Hansen teaches that "[i]n this way, the Command Processor acts as a front-end to the operating system and/or to the applications that are executing on the host computer 22" because "[a]s each new word is received, it is selectively directed to the appropriate computer resource." EX-1006, 28:21-30.

### B. Petitioners Fail to Demonstrate Motivation to Combine Taschereau with Hansen

65. In my opinion, a POSITA would not have been motivated to make Petitioners' proposed combination.

66. Petitioners and their expert, Mr. Lipoff, allege that "[a] POSITA would have been motivated to incorporate teachings concerning the Hansen speech

recognition implementation into the Taschereau speech recognition system." Pet., 10; EX-1003, ¶¶160-161. I disagree.

67. Petitioners and Mr. Lipoff offer three suggested motivations to combine in support of the proposed combination. *See* Petition, 10-11; EX-1003, ¶¶160-165.

68. ***First***, Petitioners allege that a POSITA would have "been motivated to search for references providing" details as to how an ASR system converts speech to text because Taschereau "does not disclose how" its system does so. Petition, 10; EX-1003, ¶¶160-161. Petitioners therefore propose that "[a] POSITA would have been led to Hansen which discloses well known speech recognition processing techniques." In my opinion, this does not demonstrate a motivation to combine the Taschereau and Hansen references in the manner alleged in the Petition and, in any event, would not have arrived at the claimed invention even if combined. Hansen's "processing techniques" include "immediately pass[ing]" identified words to "the 'Command Processor' portion of the algorithm" which "determines whether the word or phrase constitutes text that should be passed as data to a higher level application . . . or whether it constitutes a command." EX-1006, 10:12-19, 27:34-38. Hansen teaches that the Command Processor functionality is what allows the disclosed system to "act as a means for entering data and/or commands to a standard personal computer" in order to "essentially replace[], or supplement[] other

computer input devices, such as keyboards and pointing devices." EX-1006, 28:26-30. A POSITA would have understood that Hansen's processing techniques, and stated purpose, were therefore unsuitable for Taschereau's telephone directory assistance system.

69. Furthermore, contradicting this supposed motivation, Petitioners rely heavily on Taschereau's ASR system and even allege that Taschereau—entirely separate from Hansen—discloses the claimed "speech recognition engine." Petition, 13. Indeed, a POSITA would have disagreed with Petitioners suggestion that Taschereau "does not disclose how" its ASR functions, because Taschereau discloses over the span of 12 pages "[t]he preferred ASR system according to the invention." EX-1005, 7:24-19:4; *see also id.*, 19:10-22:23.

70. Additionally, as I explain above, Taschereau distinguishes its invention from prior art solutions that improve the efficiency of a single pass, in favor of its multi-pass process using a grammar that is "designed to be as small as possible." EX-1005, 4:20-22; 6:8-7:22. Hansen, on the other hand, teaches that its "first dictionary contains a database of the most commonly used English words." EX-1006, 24:39-56. Although Hansen discloses that "[o]ther dictionaries may be comprised of more specialized words, and may vary depending on the particular application" (EX-1006, 24:52-56), those specialized or user defined dictionaries are used by "re-executing the searching algorithm" after the first database is searched.

Taschereau, on the other hand, distinguishes prior art that uses large grammars (like Hansen's first dictionary) as well as those that employ the "common practice to split large grammars . . . into more specific and smaller grammars" (like Hansen's other dictionaries), because it will "increase the overall time associated with obtaining the desired result; and increase the interaction duration, which in turn increases costs." EX-1005, 2:20-3:3.

71. ***Second***, Petitioners allege that the combination would "improve the speech recognition system." Petition, 10-11; EX-1003, ¶162. This second alleged motivation to combine relies on Taschereau's teaching that "[t]he ASR system can also be improved through additional audio processing in addition to or in place of gain control, for example by examining and adjusting for attributes particular to the utterance to be recognized . . . ." EX-1005, 18:6-9. Petitioners rely on this disclosure to supply details for "how [Taschereau's] ASR system converts speech to text" using Hansen's speech recognition approach (Petition, 10), but in my opinion a POSITA would have understood Taschereau to be suggesting ***additional*** processing and only suggests that this additional processing could replace ***gain control***, not the entire ASR system. In my opinion, Petitioners stretch this statement by Taschereau too far when proposing that a POSITA would have reason to look to Hansen's ASR for details on how to convert speech to text using Taschereau's disclosed ASR system. Indeed, Mr. Lipoff acknowledges that Taschereau's audio processing could be

improved, but in my opinion, this is an insufficient motivation to add Hansen as a POSITA would have understood this would not have added any new functionality to Taschereau and would not have arrived at the claimed invention. *Cf.* EX-1003, ¶162.

72. ***Third***, Petitioners allege "the combination is nothing more than the application of a known technique (mapping phonemes to syllables phonemically represented in an acoustic grammar as disclosed in Hansen) to a similar known system (Taschereau's ASR system) that operates in the same way (analyzing utterances through investigation of specific words and dynamic development of grammars) for the above reasons." Petition, 11; EX-1003, ¶163.

73. For one, the two references do not operate in the same way because Taschereau distinguishes prior art "single pass" systems like Hansen from its "multi pass" system that uses grammars that are designed to be "as small as possible." EX-1005, 6:13-24. In this way, and because the grammar is designed to be small, Taschereau's system employs a "dynamically generated 'latent grammar'" because it "know[s] that the ASR system will be used again to perform one or more latent, and optionally concurrent, recognitions, each latent recognition evaluating the terms from a previous recognition process." EX-1005, 6:16-19, 14:7-8. Hansen, on the other hand, uses a first dictionary that a POSITA would have understood is a large grammar. EX-1006, 24:39-56. While Hansen's "Dictionary Lookup" function can

evaluate multiple dictionaries, it does so serially as part of a single pass because it only evaluates an additional dictionary if "no word possibilities have yet been identified" at program step 358 (with reference to Figure 10). EX-1005, 27:16-26, Fig. 10. A POSITA would have understood this to disclose a single pass, as additional dictionaries are not consulted if Hansen returns one or more than one possibilities (put another way, the additional dictionaries are used as a continuation of Hansen's single pass since the first dictionary identified 0 possibilities). Moreover, while Taschereau "dynamically develops grammars," Hansen does not and instead uses static and pre-defined dictionaries.

74. This leads to another difference between the operation of Taschereau versus Hansen. As noted above, Taschereau "whispers" utterances to human operators when it has low confidence or "the ASR system is not suitable for a response." EX-1005, 17:3-16. Hansen does not utilize such a human operator backup and instead ***discards*** sounds in the utterance which it refers to as "inactive sequences." EX-1006, 25:37-26:12 ("If [the host processor] determines, after scanning the dictionary database, that the current sequence would not ever lead to a valid word, even if additional sounds were appended, then . . . this sequence is marked 'inactive'" and after a primary word has been completed, "[t]he processor will then discard all of the inactive sequences."). Hansen also discards recognized words after its Command Processor in certain circumstances. EX-1006, 28:4-11. A

POSITA would not have been motivated to implement Hansen's practice of discarding both unrecognized and recognized words, especially given Taschereau already included a solution to this problem through other means. Indeed, Petitioners do not address these teachings in Hansen at all.

75. Additionally, I note that Petitioners' third alleged motivation relies on "the above reasons," meaning the first two alleged motivations to combine. I disagree with the first two alleged motivations to combine for the reasons discussed above, which I incorporate by reference here. Regardless, this alleged motivation to combine is flawed because, even if accurate, it does not pertain to the combination Petitioners propose.

76. Petitioners and Mr. Lipoff allege that a POSITA would have had a reasonable expectation of success in combining Taschereau and Hansen as proposed, and that the results would have been predictable, because "[c]ombining Taschereau with Hansen would merely integrate ***one component*** (mapping phonemes to syllables phonemically represented in an acoustic grammar) into Taschereau's system of receiving and responding to user's utterances." Petition, 11 (emphasis added); EX-1003, ¶164. Petitioners' alleged reasonable expectation of success is specific: the proposed combination incorporates only "one component" of Hansen. This "one component" is the alleged "mapping phonemes to syllables phonemically represented in an acoustic grammar" functionality in Hansen's Linguistic Processor.

77. But to satisfy limitations in Claim 1, Petitioners rely on Hansen's Command Processor. Petition, 25 (relying on Hansen's "'Command Processor' and associated host computer"). Therefore, Petitioners' argument that there would be a reasonable expectation of success in the combination does not apply to the combination Petitioners actually rely upon in Ground 1 (and, as a result, in Grounds 2 and 3 as well). In fact, Petitioners provide *no* explanation of why a POSITA would expect success in a combination that includes Hansen's command processor. It is my opinion that combining Hansen with Taschereau would lead to an architectural mismatch because their structures are incompatible. Petitioners never allege that a POSITA would have been motivated to modify Hansen to avoid using the Command Processor, or that a POSITA would be motivated to depart from Hansen's single pass Linguistic Processor, or that a POSITA would have had a reasonable expectation of success in making such modifications.

78. As I explain above, Hansen instructs that words and sequences can be ignored and discarded, but Petitioners or Mr. Lipoff never explain how ignored words and operating system command outputs from Hansen's single pass would function in Taschereau's multi-pass process.

79. Taschereau "requires" that the recording of an utterance be reapplied in a way where the "user does not hear the original utterance even though it is used more than once" for its multi-pass process. EX-1005, Abstract, 7:10-11, 13:8-19; *see*

*also* Petition, 22-24. Petitioners never address how Hansen's process would work where utterances are repeatedly passed through the system, including how such repeat utterances would function in the Linguistic Processor as well as the Command Processor. In my opinion, a POSITA would not have been motivated to apply Taschereau's multi-pass process to Hansen's single pass process. In addition, it is my opinion that a POSITA would not have expected success in such a combination.

80. I understand that Petitioners never argued that Hansen would be modified to avoid using the Command Processor. Petition, 24-26. Therefore, for example, each Taschereau "pass" through Hansen's Command Processor would result in operating system or application commands to the computer. It is my opinion that a POSITA would have understood that Hansen's operating system commands and application command outputs would not be suitable with Taschereau's telephonic directory assistance system.

**C. Claims [1B]/[27B]**

81. It is my understanding that Petitioners allege "Taschereau thus discloses a 'speech recognition engine'" as required by Claims [1B]/[27B][1]. Petition,

[1] Claim [1B] recites: "recognizing one or more words or phrases in the natural language utterance at a speech recognition engine coupled to the input device,

13. It is my understanding that Petitioners separately allege "Hansen thus discloses a 'speech recognition engine" as required by Claims [1B]/[27B]. Petition, 14.

82. Even if Taschereau discloses a speech recognition engine as the Petition asserts, and even if Hansen discloses a speech recognition engine as the Petition asserts, Petitioners do not explain how, in the combined system, the Taschereau-Hansen combination discloses Claims [1B]/[27B]. Likewise, Petitioners do not provide any support as to how such a combination might work. For example, Petitioners do not explain how, in the combined system both Taschereau and Hansen could provide the "speech recognition engine" that recognizes the words or phrases in the natural language utterance. Nor do Petitioners explain whether a POSITA would be motivated to make modifications to either, or both, of the alleged speech recognition engines in each of Taschereau and Hansen to make them suitable for combination. As I discuss above, which I incorporate by reference here, Hansen's processing techniques, and stated purpose, are unsuitable for Taschereau's telephone

---

wherein recognizing the words or phrases in the natural language utterance includes." Claim [27B] recites: "a speech recognition engine coupled to the input device, wherein the speech recognition engine recognizes one or more words or phrases in the natural language utterance, wherein to recognize the words or phrases in the natural language utterance, the speech recognition engine is configured to."

directory assistance system and Taschereau distinguishes approaches like Hansen's which rely on large grammars and split dictionaries. Indeed, it is my opinion that a POSITA would not have been motivated to merge the two alleged speech recognition engines nor would a POSITA have had a reasonable expectation of success in doing so.

**D. Claims [1D]-[1E]/[27D]-[27E]**

83. Petitioners' lack of clarity as to the alleged speech recognition engine is made even less clear when considering Petitioners' arguments for later limitations. For Claims [1D]/[27D][2], Petitioners rely entirely on Hansen. Petition, 18-22. Petitioners never cite to or rely upon Taschereau for Claims [1D]/[27D].

84. Specifically, Petitioners rely on Hansen's "Dictionary Lookup; "Word Collocations;" and "Grammar Check." Petition, 18-19. Hansen discloses that these

[2] Claim [1D] recites: "generating a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases." Claim [27D] recites: "generate a preliminary interpretation for the natural language utterance from the one or more syllables, wherein the preliminary interpretation generated from the one or more syllables includes the recognized words or phrases."

are part of the "series of linguistic program tools" that a POSITA would have understood make up the "Linguistic Processor." EX-1006, 10:8-11. These linguistic program tools are what Petitioners allege discloses the "speech recognition engine" of Claims [1B]/[27B] as discussed above.

85. Accordingly, Petitioners cannot demonstrate that Claims 1 or 27, as a whole, are obvious *without* Hansen's alleged speech recognition engine. This introduces problems with other limitations, which I discuss below, where Petitioners rely upon Taschereau's ASR to allegedly satisfy other limitations but nowhere explain whether or how a POSITA would have combined or modified Taschereau's ASR with Hansen's system such that features of both systems could be utilized. What is missing is an assertion that the aspects of Hansen that Petitioners require to satisfy the whole of Claims 1 and 27 would be used successfully with the aspects of Taschereau that Petitioners require be used to meet other limitations of Claims 1 and 27.

86. As Petitioners admit with respect to Claims [1E]/[27E], Hansen teaches that "[a]s words are extracted from the incoming speech signal by the Linguistic Processor, they are ***immediately*** passed to the next function in the overall program method referred to as the 'Command Processor.'" Petition, 24 (citing EX-1006, 27:34-37) (emphasis added).

87. Accordingly, Petitioners allege that Hansen's "'Command Processor' and associated host computer together constitute the claimed 'conversational language processor' because they are coupled to a speech recognition engine and interpret the recognized words or phrases." Petition, 25.

88. Thus, the proposed functionality of Hansen that Petitioners rely upon for the whole of Claims 1 and 27 includes each of: Hansen's Linguistic Processor, Command Processor, and the associated host computer.

89. But Petitioners *also* allege, separately, that Taschereau's "multi-pass latent recognition process . . . is a 'conversational language processor.'" Petition, 23. Taschereau's latent recognition process "can take advantage of linguistics (by using variations of the words in the actual listing)." EX-1005, 23:6-7. Taschereau does this by running "several latent recognition passes (perhaps using amended utterances)" and "using a subset of the large dynamic grammar" to complete several passes. EX-1005, 23:10-14.

90. A POSITA would have understood these processes are incompatible. Hansen runs its linguistic program tools to output words directly to the Command Processor, which routes commands to the computer operating system or applications. On the other hand, Taschereau uses variations of words to take advantage of linguistics and run several passes with the amended utterances. It is not clear whether Petitioners propose that multiple Taschereau passes are run *within*

Hansen's linguistic processing or whether Hansen's command outputs are to be re-applied in Taschereau's latent recognition process, or some other combination of the two, but regardless, it is my opinion that Petitioners have not demonstrated (in part because of this lack of clarity) that a POSITA would have been motivated to arrive at the claimed invention with a reasonable expectation of success.

91. Once again, like with the alleged speech recognition engine, Petitioners fail to explain how the ***combination*** satisfies the claimed conversational language processor. *See* Petition, 22-26. In my opinion, and for the reasons discussed above, a POSITA would not have been motivated to combine Hansen's Linguistic Processor, Command Processor, and the associated host computer with Taschereau and a POSITA would not have had a reasonable expectation of success in such a combination.

## VIII. REBUTTAL TO GROUND 2

92. For the reasons below, it is my opinion the Taschereau-Hansen-Franco combination does not render obvious claims 12-18, 25-26, 38-44, and 51-52.

### A. Overview of Asserted Art

#### 1. Taschereau and Hansen

93. I incorporate by reference my discussion of Taschereau and Hansen above.

**2. Franco**

94. Franco discloses improved recognition of utterances. EX-1007, Abstract.

95. Franco's solution relies on "combining acoustic evidence from the first utterance with acoustic evidence from the subsequent utterance in order to recognize the first utterance." EX-1007, Abstract. This solution "assume[s] that, if the first utterance has been incorrectly recognized on a first attempt, the user will repeat the first utterance . . .in the subsequent utterance." *Id*.

96. Petitioners acknowledge that Franco discloses an approach where "the acoustic evidence from the previous and subsequent utterances can be combined." Petition, 35.

97. Franco's assumption that the subsequent utterance is entirely or largely a repeat of the first is key. EX-1007, ¶[0006]. If that assumption did not apply, Franco's solution would combine two recordings that do not relate to each other acoustically, resulting in gibberish. *See* EX-1007, ¶¶[0017]-[0020].

98. A POSITA would have understood this limited use case, use of repetitive utterances to improve recognition, to limit the applicability of Franco. EX-1007, Abstract, ¶[0006].

### B. Petitioners Fail to Demonstrate Motivation to Combine Taschereau-Hansen with Franco

99. Petitioners provide three alleged motivations to "integrate the speech recognition error correction features disclosed in Franco with the speech recognition systems disclosed in Taschereau and Hansen." Petition, 36. In my opinion, Petitioners fail to demonstrate a motivation to combine these three references in the manner proposed.

100. ***First***, Petitioners allege that "a POSITA would understand Taschereau express[es] a motivation to combine its system with systems like Franco, that specialize in correcting errors in speech recognition systems." Petition, 36. But this logic is flawed because Franco's solution applies only in the limited use case of merging repetitive utterances to improve recognition. In addition, Petitioners' alleged motivation stems from Taschereau's recognition that "[a] significant limitation with ASR systems in the prior art is that ***as a grammar's size increases***, its accuracy diminishes . . . because as the number of possible phonetic matches for an utterance increase, the probability for error also increases as the differences between the various possible matches is smaller." EX-1005, 2:1-4 (emphasis added). A POSITA would have understood that Franco's solution does not address Taschereau's problem of large grammars. But, even if it did, Taschereau already solved the problem of large grammars by using small grammars, therefore, using

large grammars as motivation to combine Franco to help Taschereau with this problem is unjustified.

101. ***Second***, Petitioners allege motivation to combine because Franco "suggest[s] that the method of error correction . . . . was broadly applicable and could be implemented and combined with speech recognition applications, like Taschereau and Hansen." Petition, 36; EX-1003, ¶267. But, even if Franco did disclose broad applicability to other systems, a POSITA would have understood the narrow ***purpose*** for which Franco could be used—improving recognition by acoustically merging repetitive utterances—restricted Franco's ability to solve the problem of the claimed invention. Thus, in my opinion, this argument fails to provide a basis as to why a POSITA would have been motivated to modify Taschereau and Hansen in the proposed way to arrive at the claimed invention with a reasonable expectation of success.

102. ***Third***, Petitioners allege that a POSITA "would have been motivated to modify [Taschereau's] ***recursive process*** (which includes 'multiple simultaneous passes of the same utterance' (EX-1005, 19:20-20:7)) with identified 'predetermined events' like a user repeating the same utterance in order to reduce processing time." Petition, 36-37 (emphasis added). But Taschereau's recursive process is precisely what would lead a POSITA away from implementing Franco. Petitioners fail to articulate how the recursive process would be used with Franco's merging of repeat

utterances. Petitioners also do not address that Ground 2 relies on the modification of Taschereau with Hansen and therefore Petitioners have not provided any evidence that a POSITA would have been motivated to incorporate Franco into the combined Taschereau-Hansen system with a reasonable expectation of success. I disagree that a POSITA would be motivated to do so.

103. Franco assumes that repeat utterances are the same and merges them for better recognition, but Taschereau functions by repeatedly passing the utterance through its recognition system. In combination, Franco's process would assume each successive Taschereau pass is a repeat to be merged. Beyond this, Petitioners' alleged benefit, "to reduce processing time," does not follow from implementation of Franco's approach because Franco merges multiple recordings to recognize an utterance, which requires *more* processing, and Petitioners do not explain how this would "reduce processing time" in Taschereau. Hansen introduces more complexity to the proposed (but unspecified) combination. It is not clear at what point in Hansen's process—which can involve sequential dictionaries that are not adjusted applied in a single pass—that Petitioners propose Franco's merging process would be implemented. In fact, Petitioners never explain how the proposed combination of Franco and Taschereau would function at all, let alone the combination of Franco, Taschereau, and Hansen. Instead, Petitioners and their expert simply recite how Franco's system works when alleging that the '176 Patent claims would have been

obvious in Ground 2. But even if Petitioners had, the proposed combination of Franco, Taschereau, and Hansen would increase processing time, not reduce it. Accordingly, it is my opinion that Petitioners have failed to demonstrate that a POSITA would have been motivated to combine these references.

### C. Independent Claims 14 and 40 and Claims 12, 38.

104. Beyond the flawed motivation to combine, Petitioners' proposed combination also fails to disclose Independent Claims 14 and 40. Specifically, Claims [14F] and [40F],[3] which Petitioners allege each require "determining an incorrect interpretation occurred and reinterpreting" an utterance. *See* Petition, 37-43. Petitioners do not demonstrate that Franco determines that an interpretation was

[3] Claim [14F] recites: "determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event." Claim [40F] recites: "an adaptive misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event."

incorrect or that the utterance is reinterpreted. Instead, Franco performs recognition by merging repetitive utterances because Franco assumes the utterances are repeats.

105. Petitioners' descriptions of Franco's two examples are internally inconsistent. Both examples combine hypotheses of a first utterance, but while Petitioners admit the first example is "an effort to correct recognition," Petitioners describe the second example as "reinterpret[ing] the misrecognized words from the first utterance." Petition, 40, 42. Petitioners do not explain their view that the first example discloses recognition while the second discloses reinterpretation, and do not cite to any support to Franco or any other reference for that assertion.

106. Petitioners and their expert rely on analysis for Claims [14] and [40] for dependent Claims [12] and [38]. *See* Petition, 43. Therefore, for the reasons discussed above, it is my opinion that Petitioners have failed to show the asserted prior art discloses these claims.

## IX. REBUTTAL TO GROUND 3

107. As noted above, it is my opinion that Petitioners' proposed combination of Taschereau, Hansen, and Jones does not render obvious claims 3-8, 19-24, 29-34, and 45-50 of the '176 patent because Ground 3 relies on the combination of Ground 1, further modified by Jones, but does not challenge any independent claims.

108. Petitioners do not allege that Jones resolves or addresses any of the functionality relied upon in Ground 1 for the independent claims challenged therein. *See* Petition, 45-54.

109. Therefore, for the reasons discussed above for Ground 1, Petitioners do not demonstrate that Ground 3 renders any of the challenged claims obvious.

## X. REBUTTAL TO GROUND 4

110. For the reasons below, it is my opinion the Ramer-Cross combination does not render obvious claims 1-52.

### A. Overview of Asserted Art

#### 1. Ramer

111. Ramer discloses "[i]mproved mobile content presentation capabilities . . . for mobile communications facilities, such as cell phones." EX-1009, Abstract.

112. Ramer explains that "[o]nline search driven by Web-based search engines has proven to be one of the most significant uses of computer networks such as the Internet" and that "users of many mobile devices such as cell phones encounter difficulties using search technologies intended for conventional online use." EX-1009, 1:53-61. Ramer discloses that these "[d]ifficulties include the inability to display appropriate content, difficulty entering queries and taking other suitable actions such as navigation in an environment adapted to full screen displays, full-sized keyboards, and high-speed network connections." EX-1009, 1:61-65.

113. Ramer discloses that its solution addresses "[a] need [that] exists for improved search capabilities adapted for use with mobile communication devices" because "Internet search engines are currently unable to optimally deliver search results for a mobile communication facility because these search engines are specifically designed for the Internet and not mobile uses." EX-1009, 1:66-2:4.

114. Ramer thus discloses "a method and system [that] includes receiving a search request from a mobile communication facility, receiving information relating to a user characteristic associated with the mobile communication facility, and presenting search results prioritized according to relevancy, wherein the relevancy is at least in part based on the information relating to the user characteristic associated with the mobile communication facility" and "[t]he prioritized search results may be presented in a prioritized list according to the search results relevancies." EX-1009, 2:8-17.

115. Ramer teaches that these search requests can take two different forms, "[a] content request may be made by an explicit query requested entered by a mobile communication facility user ***as text in the query facility*** of the mobile communication facility" or "[a] content request may be made by an explicit query request entered by a mobile communication facility user ***as a voice command in the voice recognition query facility*** of a mobile communication facility." EX-1009, 2:18-26 (emphasis added).

116. Certain embodiments of Ramer concern queries entered as text in the query facility, while other embodiments concern queries entered as voice in the voice recognition query facility. EX-1009, 5:49-51 ("The mobile communication facility 102 may facilitate the collection of data from data sources as a result of a query entry 120 or voice entry 122."); *see also* EX-1009, 3:6-53, 4:2-19, 5:48-63, 6:17-26, 24:63-25:34.

117. Ramer discloses that typed query entry "may be accomplished through the use of a numeric key pad entry, full mobile device keyboard entry (e.g., that found on a Blackberry or Treo device), partial mobile device keyboard entry ( e.g., that found on a Blackberry device with only one key for every two letters), stylus/handwriting entry, bar code scanner ( either 2D bar code or 3D bar code: "Quick Response Code"), or photographic entry using cellular phone-camera; through other navigational facilities ( e.g., a stylus, arrow keys, scroll wheel, etc.); or through access to a computer network, such as through a physical connection (e.g., Ethernet or other network cable, wire, or the like), or through infrared, RF, Bluetooth or other wireless query entry." EX-1009, 5:48-63.

118. Ramer discloses that "[t]he voice entry 122 function of the mobile communication facility may be used through the speaker-receiver device of the mobile communication facility 102 or by use of the standard SMS lexicon and syntax" and "[t]he voice entry 122 function may permit voice dialing, voice memo,

voice recognition, speech recognition, or other functions related to audible input." EX-1009, 6:17-26.

**2. Cross**

119. Cross teaches that "[c]urrent lightweight voice solutions require a developer to build a grammar and lexicon to limit the potential number of words that an ASR engine must recognize-as a means for increasing accuracy." Cross also explains that "speaker independent voice recognition is implemented to enhance the user experience and interaction" with devices and "[t]he state of the art in speaker independent recognition allows for some sophisticated voice applications to be written as long as there is a limited vocabulary associated with each potential voice command." EX-1010, 1:51-67.

120. Cross receives speech and passes the speech along to an ASR engine which is "a module of computer program instructions, also stored in RAM in this example, that includes an ASR lexicon (106) of words capable of recognition by the ASR engine," which Cross discloses "is an association of words in text form with phonemes representing pronunciations of each word." EX-1010, 6:26-32.

**B. Independent Claims [1C]/[27C].**

121. At the outset, I note that Petitioners and their expert allege that Cross discloses Claims [1C]/[27C]. Petitioners exclusively rely on Cross to meet these limitations and do not cite to Ramer. Petition, 60-61.

122. It is my understanding that Petitioners interpret the term "acoustic grammar" in Claims [1C]/[27C] to require "a grammar of phototactic [sic] English language rules that maps phonemes to syllables" or "a collection of the phonemes linked together to form syllables, which are linked together to form the words of the language." Petition, 60-61. For purposes of this Declaration in support of Patent Owner's Preliminary Response, I offer no independent opinion of the proper interpretation of Claims [1C]/[27C] but have adopted the interpretation proposed by Petitioners. I reserve the right to further address the interpretation of limitation [1C]/[27C] if this proceeding is instituted.

123. Petitioners allege that each of these constructions are satisfied by Cross's disclosure of a table with "phonemes representing, to the extent that it is practically feasible to do so, all pronunciations of all the words in a particular language." Petition, 61 (citing EX-1010, 6:38-43). Cross discloses a table of words, specifically, "lexicon (106) is an association of ***words in text form*** with phonemes representing pronunciations of each word." EX-1010, 6:31-33 (emphasis added).

124. Petitioners acknowledge that Cross addresses pronunciations, not syllables, but nevertheless allege that "[a] POSITA would understand the reference to 'pronunciations' in Cross to be synonymous with syllables." Petition, 61 (citing EX-1003, ¶¶411-412). In my opinion, Petitioners fail to establish this assertion.

125. Petitioners' expert does not provide any support for the assertion that a POSITA would have understood pronunciations are synonymous with syllables. EX-1003, ¶412.

126. In fact, it is unclear what Petitioners' expert means by the assertion that "pronunciations [are] synonymous with syllables." To the extent he means that pronunciations and syllables are the same thing, I disagree. For example, the words "either" and "either" have two pronunciations but the same number of syllables. In contrast, "chocolate" and "chocolate" are pronounced differently where one pronunciation has two syllables while the other has three (or "learned" and "learned," with one or two syllables respectively).

127. Petitioners and their expert cite to references (Verma[4] and Jones '330[5]) which are not asserted in Ground 4 (or any ground) to allege that mapping phonemes to syllables was "well known in the art." Petition, 61. Mr. Lipoff asserts the same, citing to his analysis in Ground 1. *See* EX-1003, ¶412.

128. Mr. Lipoff, in Ground 1, characterizes Verma as disclosing "phonemes of speech utterances being mapped to the syllables by using a syllabic mapping

[4] U.S. Pat. Publ. No. 2007/0067174 (Exhibit 1016).

[5] U.S. Pat. No. 6,804,330 (Exhibit 1017).

database." EX-1003, ¶199. He further cites Verma's disclosure that "[t]he syllabic mapping database maps groups of phones or phonemes to syllables, so that the phones or phonemes of the speech utterance that have been identified can be grouped together into syllables. In this way, by first segmenting the phones or phonemes of a speech utterance and then grouping sequences of the phones or phonemes into syllables, the syllables of the speech utterance are identified." *Id*. (citing EX-1016, ¶[0025]).

129. Also in Ground 1, Mr. Lipoff addresses the "Nuance 7.0 speech server" that is referenced in Jones '330. *See* EX-1003, ¶200 (citing EX-1017, 7:57-65; 8:56-61). Mr. Lipoff alleges that "[a]s individual phonemes are identified, they are chained together to form syllables . . . the syllables will then be chained together [and] [w]hen applicable, words are combined into phrases." *Id*. (citing EX-1017, 8:27-31).

130. In my opinion, these disclosures support my opinion that Petitioners have failed to show the claims are obvious in view of Cross. Petitioners allege these unasserted references establish it was known to map phonemes to syllables, yet also acknowledge that Cross disclosed pronunciations, not syllables. Indeed, Petitioners never allege that Cross explicitly discloses mapping phonemes to syllables.

131. In my opinion, whether Verma and Jones '330 may suggest certain functionality does not mean that a POSITA would have understood Cross to disclose

that functionality; yet that is what Petitioners and Mr. Lipoff seem to argue. I disagree that these separate references would have demonstrated to a POSITA that the combination of Ramer and Cross discloses Petitioners' interpretations of "acoustic grammar."

132. Moreover, Petitioners and Mr. Lipoff do not explain (or even suggest) how these references would be used to modify Cross. For example, Mr. Lipoff never suggests that Cross would (or could) apply Verma's "syllabic mapping database" or that a POSITA would have expected success in such a modification. Likewise, Mr. Lipoff does not offer the opinion that Cross would (or could) utilize the Nuance 7.0 speech server instead of its ASR lexicon.

133. Accordingly, it is my opinion that Petitioners have not demonstrated that Ground 4 satisfies the interpretations that Petitioners allege should be applied for "acoustic grammar" (again, without offering an independent opinion of the proper interpretation of this term).

### C. Independent Claims 14 and 40.

134. Petitioners allege that Ramer discloses the "adaptive misrecognition engine" claimed in independent Claims 14 and 40.[6] Petition, 66-68. Petitioners exclusively rely on Ramer to meet these limitations. Petition, 67.

135. Petitioners allege that Ramer discloses a scenario where "the user ***uttered*** a word that did not appear in its dataset." Petition, 67-68 (citing EX-1009, 25:9-25) (emphasis added). I disagree. Ramer explicitly states that this scenario utilizes ***text*** that the user inputs into "query entry 120." EX-1009, 25:10-25, Fig. 4, 9; *see also* 5:48-67 ("Query entry 120 may be accomplished through the use of a numeric key pad entry, full mobile device keyboard entry . . . .").

[6] Claim [14F] recites: "determining that the conversational language processor incorrectly interpreted the words or phrases in response to an adaptive misrecognition engine detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event." Claim [40F] recites: "an adaptive misrecognition engine configured to determine that the conversational language incorrectly interpreted the words or phrases in response to detecting a predetermined event, wherein the conversational language processor reinterprets the words or phrases in response to the predetermined event."

136. Ramer distinguishes "query entry 120" from "voice entry 122." EX-1009, 5:48-67. A POSITA would have understood that query entry 120 is not the same as voice entry 122. Accordingly, when Ramer refers to embodiments that utilize inputs from query entry 120, a POSITA would have understood such embodiments to address written (e.g., typed) inputs rather than natural language utterances.

137. The '176 Patent describes "existing systems [which] tend to have complex human to machine interfaces" and "many existing voice user interfaces (when they actually work) still require significant learning on the part of the user." EX-1001, 1:33-35. The '176 Patent solves these and "other problems" that "[e]xisting systems suffer from." EX-1001, 1:62. Accordingly, the invention of the '176 Patent "relates to selecting and presenting advertisements based on natural language processing of voice-based inputs" where "[t]he voice-based input may include a user utterance, and a request may be identified from the utterance." EX-1001, 1:8-10, 2:3-5.

138. A POSITA would not have been motivated to combine or modify Ramer's "query entry" embodiments to arrive at the claimed invention of the '176 patent to solve the problem of selecting and presenting advertisements based on natural language processing of ***voice-based*** inputs. Petitioners never acknowledge the difference.

139. In fact, far from acknowledging that the relied-upon passages of Ramer address typed-entries rather than spoken utterances, Petitioners ignore it. Petitioners argue that when "a user ***utters*** 'Coltrain' . . . [t]he system may misrecognize the word." Petition, 68 (emphasis added); (citing EX-1009, 25:9-25).

140. In my opinion, a POSITA would have understood that the user doesn't utter Coltrain in Ramer's example, they type it into "query entry 120." It therefore follows that the system would not "misrecognize" what the user typed because the system would "recognize" exactly what was typed. Therefore, in my opinion, Petitioners are wrong to suggest that Ramer's "[s]ystem may misrecognize the word as 'Coal Train' or 'Soul Train.'" Petition, 68 (citing EX-1009, 25:9-15).

141. In another example. Figure 9 demonstrates Ramer's system with typed entry "Br Sp." Much like in the example where "Coltrane" is suggested as opposed to "Coal Train" or "Soul Train," here Ramer shows entries for "Britney Spears" first as opposed to other options like "Alicia Bridges" or "British Phone." *See* EX-1009, 12:46-49, Fig. 9. But in this example too, a POSITA would have understood that the Ramer system recognized "Br Sp" as the typed entry.

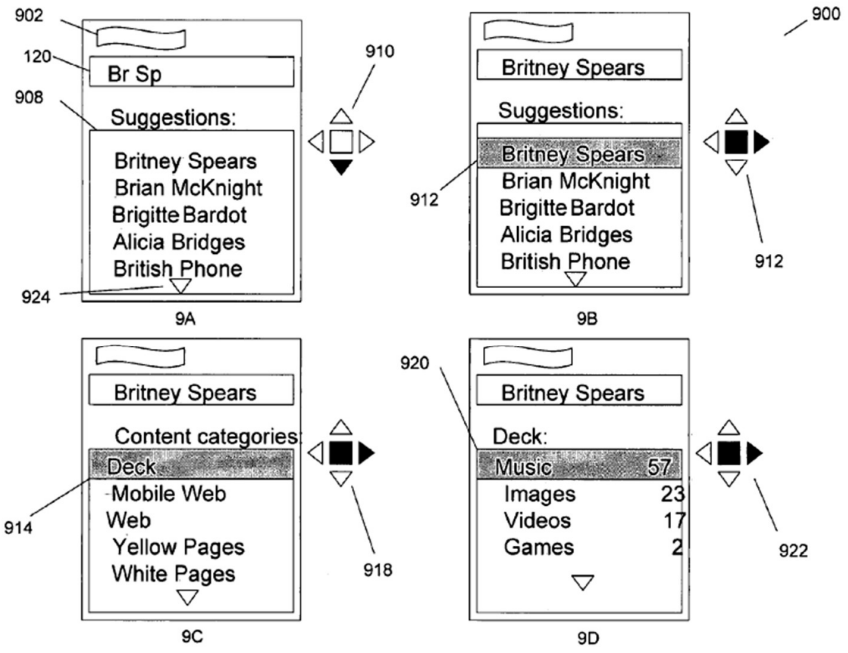


Fig. 9

142. Moreover, a POSITA would have understood that in both Ramer's "Coltrane" example and the "Britney Spears" example, alternatives are provided as contrast. Indeed, Ramer explains in the Coltrane example that "Coltrane" is suggested "as opposed to 'Coal Train' or 'Soul Train.'" EX-1009, 25:9-15. Therefore, I disagree with Petitioners' allegation that this example in Ramer demonstrates that the system "misrecognized" the user entry as "Coal Train" or 'Soul Train'" or that this example "reinterprets the words or phrases" uttered by the user. Petition, 68.

143. Petitioners next turn to another example where Ramer discloses, in Petitioners' words, "that the user uttered a word that did not appear in [Ramer's]

dataset." Petition, 67 (citing EX-1009, 25:20-25 ("[I]n those instances where 'a user's query entry 120 returns a null result set or an improbable result set,' 'the search facility . . . could automatically trigger correction 244 and iteratively cycle through alternative query entries 120 until a non-null or higher probability result set is delivered.'")). However, I disagree that this example demonstrates that Ramer's system "determines that the conversational language processor incorrectly interpreted the words or phrases."

144. For one, as noted above, query entries 120 are typed, not uttered. This matters because the claims require that the "incorrectly interpreted the words or phrases" are the "one or more words or phrases ***in the natural language utterance***" that were recognized "at a speech recognition engine coupled to the input device." Claim [14B]; *see also* Claim [40B]. Petitioners have not demonstrated that this example of Ramer discloses functionality that applies to a natural language utterance or pointed to any disclosure in Ramer where similar functionality is applied for a spoken utterance.

145. But beyond this, a POSITA would have understood that in this example, when a user's typed query entry does not appear in the dataset, the system did not "incorrectly interpret" what words or phrases were typed. To the contrary, the user simply entered a word or phrase that didn't match the dataset, Ramer never suggests that the system failed to accurately capture the typed input. For example, if the user

were to type "abcde" and Ramer's system searched for "abdce" in its dataset but returned a null result set or an improbable result set, a POSITA would not conclude that Ramer "incorrectly interpreted" the typed entry, it searched for exactly what the user typed.

146. Petitioners do not identify any other disclosure in Ramer that allegedly satisfies this limitation. Accordingly, it is my opinion that Petitioners have failed to demonstrate that Claims 14 and 40 would have been obvious.

## XI. CONCLUSION

147. In sum, it is my opinion that Grounds 1-4 of the Petition do not render Claims 1-52 of the '176 Patent unpatentable as obvious for the reasons I discuss above.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on the 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 August 20, 2025 in New York, NY.

Dr. Anatole Gershman