

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM SERVICES LLC,

Petitioner,

v.

VB ASSETS, LLC,

Patent Owner.

Case: IPR2025-01343

U.S. Patent No. 9,502,025

DECLARATION OF PADHRAIC SMYTH, PH.D.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	14
II. BACKGROUND AND QUALIFICATIONS	14
III. MATERIALS CONSIDERED	18
IV. BASIS OF OPINION	22
A. Legal Standard for Anticipation	22
B. Legal Standard for Obviousness	23
C. Level of Ordinary Skill in the Art	24
D. Claim Construction	26
V. OVERVIEW OF THE '025 PATENT	27
VI. PROSECUTION HISTORY	29
VII. RELATED LITIGATION	34
VIII. PRIOR ART TO THE CHALLENGED PATENT	34
A. Technical Background	34
B. Summary of the Prior Art	43
1. Weider (EX1004)	43
2. Gold (EX1005)	45
3. Martinez (EX1006)	47
4. Wang (EX1007)	49
5. Kennewick (EX1008)	51
6. Ginter (EX1009)	53
IX. SUMMARY OF INVALIDITY GROUNDS	54

TABLE OF CONTENTS

(Continued)

	Page
X. Invalidity Grounds	55
A. Ground 1: Weider in combination with Gold renders obvious claims 1, 3, 6-8, 10, 13-14, and 19	55
1. Claim 1	55
[1.pre] “A system for providing a natural language content dedication service, comprising:”	55
[1.a]: “one or more processors; and;”	62
[1.b] “one or more non-transitory computer readable mediums storing executable instructions that when executed by the one or more processors cause the one or more processors to:”	64
[1.b(i)] “receive a first utterance that includes a natural language utterance;”	67
[1.b(ii)] “determine, based on processing of the first utterance by a speech recognition engine, one or more words or phrases of the first utterance;”	69
[1.b(iii)] “provide the one or more words or phrases as an input to a conversational language processor;”	71
[1.b(iv)] “interpret the first utterance, at the conversational language processor, based on the one or more words or phrases;”	73
[1.b(v)] “identify, based on the interpretation of the first utterance, content to dedicate to a recipient;”	74
[1.b(vi)] “initiate, based on the identified content, a dedication to the recipient;”	76

TABLE OF CONTENTS (Continued)

	Page
[1.b(vii)] “receive a second utterance to be associated with the dedication;”	78
[1.b(viii)] “determine, based on a processing of the second utterance by the speech recognition engine, one or more words or phrases of the second utterance;”	80
[1.b(ix)] “provide the one or more words or phrases of the second utterance as textual annotations within metadata of the content; and”	81
[1.b(x)] “send information to enable the recipient to access the content and the second utterance.”	84
2. Claim 3: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to annotate the content with the second utterance.”	88
3. Claim 6	88
[6.a] “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: send, to the recipient, an electronic link to access the content as the information to enable the recipient to access the content; and”	88
[6.b] “send, to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance.”	90
4. Claim 7: “The system of claim 1, wherein the content includes one or more of a song or a video.”	93
5. Claim 8	93

TABLE OF CONTENTS

(Continued)

	Page
6. Claim 10: “The method of claim 8, further comprising annotating the content with the second utterance.”	95
7. Claim 13: “The method of claim 8, further comprising: sending, to the recipient, an electronic link to access the content as the information to enable the recipient to access the content; and sending, to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance.”	95
8. Claim 14: “The method of claim 8, wherein the content includes one or more of a song or a video.”	95
9. Claim 19: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide the second utterance within the content.”	95
B. Ground 2: Weider in view of Gold and Martinez renders obvious claims 2, 9, 17, and 18	96
1. Claim 2: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide an image as one or more tags within metadata of the content.”	96
2. Claim 9: “The method of claim 8, further comprising providing an image as one or more tags within metadata of the content.”	101
3. Claim 17: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide an electronic link associated with the content as one or more tags within metadata of the content.”	101

TABLE OF CONTENTS (Continued)

	Page
4. Claim 18: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide a voice tag as one or more tags within metadata of the content.”	103
C. Ground 3: Weider in view of Gold and Ginter renders obvious claims 4 and 11	104
1. Claim 4: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to purchase the content from a content provider to enable the recipient to access the content.”	104
2. Claim 11: “The method of claim 8, further comprising purchasing the content from a content provider to enable the recipient to access the content.”	108
D. Ground 4: Weider in view of Gold, Ginter, and Wang renders obvious claims 5 and 12	108
1. Claim 5	108
[5.a] “The system of claim 4, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: identify, for the dedication, a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option; and”	108
[5.b] “bill a user associated with the first utterance based on the purchase option.”	113

TABLE OF CONTENTS (Continued)

	Page
2. Claim 12: “The method of claim 11, further comprising: identifying, for the dedication, a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option; and billing a user associated with the first utterance based on the purchase option.”	115
E. Ground 5: Weider in view of Gold and Wang renders obvious claims 15-16	115
1. Claim 15	115
[15.a] “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: determine that a user is interacting with a media item; and”	115
[15.b] “identify, based on the natural language processing and the interaction of the user with the media item, the media item as the content to dedicate to the recipient.”	118
2. Claim 16	121
[16.a] “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: determine a media item that a user is playing;”	121
[16.b] “determine that the user provided the first utterance during the playing of the media item by the user; and”	122

TABLE OF CONTENTS

(Continued)

	Page
[16.c] “identify, based on the natural language processing and the determination that the user provided the first utterance during the playing of the media item, the media item as the content to dedicate to the recipient.”	124
F. Ground 6: Kennewick in view of Gold and Weider renders obvious claims 1, 3, 6-8, 10, 13-14, and 19	124
1. Claim 1	124
[1.pre] “A system for providing a natural language content dedication service, comprising:”	124
[1.a]: “one or more processors; and;”	129
[1.b] “one or more non-transitory computer readable mediums storing executable instructions that when executed by the one or more processors cause the one or more processors to:”	131
[1.b(i)] “receive a first utterance that includes a natural language utterance;”	133
[1.b(ii)] “determine, based on processing of the first utterance by a speech recognition engine, one or more words or phrases of the first utterance;”	135
[1.b(iii)] “provide the one or more words or phrases as an input to a conversational language processor;”	137
[1.b(iv)] “interpret the first utterance, at the conversational language processor, based on the one or more words or phrases;”	140
[1.b(v)] “identify, based on the interpretation of the first utterance, content to dedicate to a recipient;”	141

TABLE OF CONTENTS (Continued)

	Page
[1.b(vi)] “initiate, based on the identified content, a dedication to the recipient;”	143
[1.b(vii)] “receive a second utterance to be associated with the dedication;”	145
[1.b(viii)] “determine, based on a processing of the second utterance by the speech recognition engine, one or more words or phrases of the second utterance;”	147
[1.b(ix)] “provide the one or more words or phrases of the second utterance as textual annotations within metadata of the content; and”	148
[1.b(x)] “send information to enable the recipient to access the content and the second utterance.”	150
2. Claim 3: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to annotate the content with the second utterance.”	152
3. Claim 6	152
[6.a] “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: send, to the recipient, an electronic link to access the content as the information to enable the recipient to access the content; and”	152
[6.b] “send, to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance.”	152
4. Claim 7: “The system of claim 1, wherein the content includes one or more of a song or a video.”	153

TABLE OF CONTENTS (Continued)

	Page
5. Claim 8	153
6. Claim 10: “The method of claim 8, further comprising annotating the content with the second utterance.”	154
7. Claim 13: “The method of claim 8, further comprising: sending, to the recipient, an electronic link to access the content as the information to enable the recipient to access the content; and sending, to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance.”	155
8. Claim 14: “The method of claim 8, wherein the content includes one or more of a song or a video.”	155
9. Claim 19: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide the second utterance within the content.”	155
G. Ground 7: Kennewick in view of Gold, Weider, and Martinez renders obvious claims 2, 9, 17, and 18.....	156
1. Claim 2: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide an image as one or more tags within metadata of the content.”	156
2. Claim 9: “The method of claim 8, further comprising providing an image as one or more tags within metadata of the content.”	157
3. Claim 17: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide an electronic link associated with the content as one or more tags within metadata of the content.”	158

TABLE OF CONTENTS (Continued)

	Page
4. Claim 18: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide a voice tag as one or more tags within metadata of the content.”	159
H. Ground 8: Kennewick in view of Gold, Weider, and Ginter renders obvious claims 4 and 11	159
1. Claim 4: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to purchase the content from a content provider to enable the recipient to access the content.”	159
2. Claim 11: “The method of claim 8, further comprising purchasing the content from a content provider to enable the recipient to access the content.”	162
I. Ground 9: Kennewick in view of Gold, Weider, Ginter, and Wang renders obvious claims 5 and 12.....	162
1. Claim 5	162
[5.a] “The system of claim 4, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: identify, for the dedication, a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option; and”	162
[5.b] “bill a user associated with the first utterance based on the purchase option.”	165

TABLE OF CONTENTS

(Continued)

	Page
2. Claim 12: “The method of claim 11, further comprising: identifying, for the dedication, a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option; and billing a user associated with the first utterance based on the purchase option.”	166
J. Ground 10: Kennewick in view of Gold, Weider, and Wang renders obvious claims 15-16.....	166
1. Claim 15	166
[15.a] “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: determine that a user is interacting with a media item; and”	166
[15.b] “identify, based on the natural language processing and the interaction of the user with the media item, the media item as the content to dedicate to the recipient.”	166
2. Claim 16	169
[16.a] “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: determine a media item that a user is playing;”	169
[16.b] “determine that the user provided the first utterance during the playing of the media item by the user; and”	169

TABLE OF CONTENTS
(Continued)

	Page
[16.c] “identify, based on the natural language processing and the determination that the user provided the first utterance during the playing of the media item, the media item as the content to dedicate to the recipient.”	170
XI. CONCLUSION.....	171

I, Padhraic Smyth, Ph.D., declare as follows:

I. INTRODUCTION

1. I have been asked by the party requesting this review, Amazon.com Services LLC (“Petitioner”), to provide my expert opinions in support of the petition for *inter partes* review of U.S. Patent No. 9,502,025 (the “’025 patent” or “challenged patent”) challenging the patentability of claims 1-19 (the “challenged claims”).

2. In summary, it is my opinion that all challenged claims are unpatentable for the reasons I provide in this declaration.

3. I am being compensated for my time at my normal consulting rate of \$800 per hour. This compensation is not contingent upon my performance, the conclusions I reach in my analysis, the outcome of this matter, or any issues involved in or related to this matter.

II. BACKGROUND AND QUALIFICATIONS

4. The details of my education, work experience, research, patents, and publications are summarized in my CV (EX1023). I provide here a summary of some of the relevant information from my CV, as well as other relevant background information.

5. I earned a Ph.D. and Master’s degree in Electrical Engineering from the California Institute of Technology. I earned a Bachelor of Engineering (Electronic)

from the National University of Ireland, University College Galway in 1984. My senior undergraduate thesis was on speech recognition.

6. At present, I am a Distinguished Professor and Endowed Chair in Artificial Intelligence in the Department of Computer Science at the University of California, Irvine. Prior to that, I was a Member of Technical Staff and Technical Group Leader at the Jet Propulsion Laboratory.

7. As a professor at UC Irvine since 1996, I conduct research in the areas of machine learning, artificial intelligence, signal processing, pattern recognition, and natural language processing, and I regularly teach courses related to these topics. I have published over 230 peer-reviewed papers and two books on these topics and my work has been cited over 70,000 times in the research literature. My research work is funded by federal agencies such as NSF, NIH, NIST, IARPA, DARPA, ONR, and NASA.

8. During my career I have worked extensively on research related to key aspects of automated speech recognition, natural language processing, user modeling, and human dialog. I have attached my CV (EX1023), which provides more details about the experience I summarize here. As an example, since 1992 I have conducted a significant body of research and published a series of papers on the topic of hidden Markov models (HMMs). HMMs are a general statistical framework for pattern recognition in sequential data (such as audio waveforms or

text) and were a key component in automated speech recognition systems and natural language understanding since the 1980's. My research on HMMs produced significant advances in the understanding of the statistical and algorithmic foundations of speech recognition and natural language understanding. These advances included development of new HMM methods for detecting anomalous patterns in time-series signals, new algorithms for automated clustering of waveforms using HMMs, new classes of HMMs with applications to speech modeling, and novel HMM methods for automated segmentation of waveforms. My papers on this topic have been collectively cited over 1,800 times in the literature.

9. I have also conducted extensive research in the field of natural language understanding since 1999. My work included co-authoring a textbook, *Modeling the Internet and the Web: Probabilistic Methods and Algorithms*, published in 2003 with chapters devoted to algorithms for natural language understanding. I've also written papers on topics like analysis of character sequences in non-English languages, automated algorithms for extracting human-interpretable summaries of content from large text corpora, methods for automated characterization of authors from the text they generate, automated techniques for creating network and graph models of relationships between documents and entities mentioned in the documents, algorithms for analyzing text with the assistance of human-generated semantic ontologies, and analysis of large language models. As with my work on

HMMs, my work in natural language understanding meant that I was, and am, broadly familiar with the research literature in this area.

10. I have also conducted research in user modeling, focusing on statistical and machine learning techniques for automated generation of user models from observational data. In addition, I have published papers on techniques for automated analysis of human dialog, such as automatically classifying utterances into specific topics. As I discuss below, user modeling is an important function of many interactive conversational systems, which enables such systems to manage a dialog with the user and infer information about the user's requests.

11. In addition to my research and publications I have taught and lectured extensively on these topics. Since my appointment as a professor at UC Irvine in 1996, I have taught both undergraduate and graduate courses there in machine learning, pattern recognition, natural language understanding, artificial intelligence, and user modeling. In addition, I have given keynote lectures and tutorials on these topics at a variety of international conferences since the late 1990s.

12. I have also served in editorial roles and conference chair positions for major journals and conferences in my research areas, including serving as a founding editorial board member for the Journal of Machine Learning Research, an associate editor for the Journal of the American Statistical Association, program chair for the Association for Computing Machinery's Special Interest Group on Knowledge

Discovery and Data Mining conference, co-program chair for the International Joint Conference on Artificial Intelligence, and serving as Senior Area Chair for the Neural Information Processing Systems conference and the International Machine Learning Conference on multiple occasions.

13. I am also a technical consultant with private industry in areas including machine learning, user modeling, signal processing, pattern recognition, and natural language understanding. Over the past 25 years, I have consulted in these areas for companies including Toshiba, Microsoft, Google, SAP, Netflix, eBay, AT&T, Samsung, Nokia, as well as several startup companies.

III. MATERIALS CONSIDERED

14. I have reviewed and considered, in preparing this declaration, the '025 patent (EX1001), its prosecution history (EX1003), and the following prior art references:

- U.S. Publication No. 2007/0050191 to Weider et al., filed August 29, 2005 and titled “Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions” (“Weider”) (EX1004);
- U.S. Publication No. 2002/0032752 to Gold et al., filed June 7, 2001 and titled “Method and System for Electronic Song Dedication” (“Gold”) (EX1005);

- U.S. Publication No. 2007/0083762 to Martinez, filed October 10, 2005 and titled “Set of Metadata for Association with a Composite Media Item and Tool for Creating Such Set of Metadata” (“Martinez”) (EX1006);
- U.S. Publication No. 2004/0199387 to Wang et al., filed April 26, 2004 and titled “Method and System for Purchasing Pre-Recorded Music” (“Wang”) (EX1007);
- U.S. Publication No. 2004/0193420 to Kennewick et al., filed July 15, 2003 and titled “Mobile Systems and Methods for Responding to Natural Language Speech Utterance” (“Kennewick”) (EX1008);
- U.S. Patent No. 5,892,900 to Ginter et al. filed August 30, 1996 and titled “Systems and Methods for Secure Transaction Management and Electronic Rights Protection” (“Ginter”) (EX1009);

15. I am also aware of, and considered in forming my opinions presented in this declaration, other information available to a person of ordinary skill in the art at the time of the priority date of the challenged patent (“POSITA”), such as other

patents, reference materials, articles, and standards, and knowledge of a POSITA.¹

These include:

- Asthana, A. & Krzyzanowski, P., “A Small Domain Communications System for Personalized Shopping Assistance,” Proceedings of 1994 International Conference on Personal Wireless Communications (pp. 199-203), IEEE Press (1994) (“Asthana”) (EX1010);
- Lucente, M., “Conversational Interfaces for E-Commerce Applications,” Communications of the ACM (Sept. 2000), 43(9), 59-61, (“Lucente”) (EX1011);
- Huang et al., “Spoken Language Processing: A Guide to Theory, Algorithm, and System Development,” Prentice Hall (2001) (“Huang”) (EX1012);
- U.S. Patent No. 6,859,776, issued February 22, 2005 to Michael H. Cohen and titled “Method and Apparatus for Optimizing a Spoken Dialog Between a Person and a Machine” (“Cohen”) (EX1013);
- McCandless, M. et al., “The MP3 Revolution,” IEEE Intelligent Systems (May/June 1999) (“McCandless”) (EX1014);

¹ I understand that the challenged patent has a priority date of November 10, 2009, as I discuss in more detail below.

- Haughey, M., “Napster A New Killer Internet App,” *Evolt* (Nov. 12, 1999) (“Haughey”) (EX1015);
- U.S. Patent Publication No. 2005/0021478, published January 27, 2005, and titled “Method and System for Network-Based Purchase and Distribution of Media” (“Gautier”) (EX1016);
- Press Release, Apple Inc., “Apple Launches the iTunes Music Store,” (April 28, 2003) (“Apple Press Release”) (EX1017);
- U.S. Patent Publication No. 2008/0301241, published December 4, 2008, and titled “System and Method of Generating a Media Item Recommendation Message with Recommender Presence Information” (“Svendsen”) (EX1018);
- U.S. Patent Publication No. 2005/0240494, published October 27, 2005, and titled “Method and System for Sharing Playlists” (“Cue”) (EX1019);
- U.S. Patent Publication No. 2009/0069912, published March 12, 2009, and titled “Digital Media Player and Method for Facilitating Music Recommendation” (“Stefik”) (EX1020);
- Claim Construction Memorandum Order, *VB Assets, LLC, v. Amazon.com, Inc., et al.*, 1:19-cv-01410-MN (D. Del. June 23, 2021), Dkt. 90 (EX1021);

- Complaint for Patent Infringement, *VB Assets, LLC, v. Amazon.com Services, LLC*, 1:24-cv-00839-MN (D. Del. July 18, 2024), Dkt. 1 (EX1022); and
- “Microsoft Computer Dictionary,” Microsoft Press (2002) (EX1024).

16. Should new information be disclosed or discovered during this proceeding, I may supplement and elaborate on the opinions I express in this declaration as necessary. Nothing in this declaration should be interpreted as an opinion that no grounds of invalidity exist other than the ones in the petition. I do not express such an opinion and have not reached such an opinion.

IV. BASIS OF OPINION

A. Legal Standard for Anticipation

17. I understand that a claimed invention is unpatentable as “anticipated” by a single prior art reference if it discloses each and every limitation of the claim. I also understand that, even if a claim is not explicitly disclosed in the prior art reference, the prior art reference may nevertheless anticipate the claim if the subject matter of the claim is inherently present from the perspective of a POSITA. Moreover, I understand that even if the prior art reference does not disclose the limitations of a claim in the same arrangement or combination of the limitations of the claim, it may nevertheless anticipate if a POSITA would envisage the claimed arrangement or combination.

B. Legal Standard for Obviousness

18. I understand that a claimed invention is unpatentable as “obvious” if the differences between the invention and the prior art are such that the subject matter of the claim as a whole would have been obvious to a POSITA.

19. It is my understanding that “obviousness” is a question of law based on underlying factual issues including (1) the scope and content of the prior art, (2) the differences between the prior art and the asserted claims, (3) the level of ordinary skill in the pertinent art, and (4) the existence of secondary considerations such as commercial success, long-felt but unresolved needs, and failure of others.

20. I understand that a single reference can render a patent claim obvious if any differences between that reference and the claim would have been obvious to a POSITA. Alternatively, the teachings of two or more references may be combined in the same way as disclosed in the claims, if such a combination would have been obvious to a POSITA. In determining whether a combination based on either a single reference or multiple references would have been obvious, I understand that it is appropriate to consider the following factors:

- Whether the teachings of the prior art references disclose known concepts combined in familiar ways, and when combined, would yield predictable results;

- Whether a POSITA could implement a predictable variation, and would see the benefit of doing so;
- Whether a POSITA would have recognized a reason to combine known elements in the manner described in the claim;
- Whether there is some teaching or suggestion in the prior art to make the modification or combination of elements claimed in the patent; and
- Whether the innovation applies a known technique that had been used to improve a similar device or method in a similar way.

21. In an obviousness evaluation based on a combination of multiple prior art references, I understand that the prior art references themselves may provide a suggestion, motivation, or reason to combine, but other times the nexus linking two or more prior art references is simple common sense. I further understand that the obviousness analysis recognizes that market demand, rather than scientific literature, often drives innovation, and that a motivation to combine references may be supplied by the direction of the marketplace.

C. Level of Ordinary Skill in the Art

22. I understand that the '025 patent should be interpreted from the perspective of a person having “ordinary skill in the art” as of the priority date of the patent application (i.e., a POSITA). For the purposes of this declaration, I assume that the priority date of the patent application leading to the '025 patent is no earlier

than November 10, 2009, and I understand that Patent Owner has not claimed an earlier priority date in the parallel litigation. I am familiar with the technological field at issue and the state of the art as of that date.

23. I understand that the education level of those in the technological field, sophistication of the technology, types of problems encountered, prior art solutions to those problems, and the speed at which innovations are made in the field may factor into establishing what the appropriate level of ordinary skill should be.

24. The specification of the challenged patent states that it relates to “providing a natural language content dedication service in a voice services environment.” EX1001, Abstract, 1:16-18. The claims of the challenged patent are directed to a system and a method for “providing a natural language content dedication service” involving “receiv[ing] a first utterance that includes a natural language utterance,” “determin[ing], based on processing of the first utterance by a speech recognition engine, one or more words or phrases of the first utterance,” “provid[ing] the one or more words or phrases as an input to a conversational language processor,” “interpret[ing] the first utterance, at the conversational language processor, based on the one or more words or phrases,” “identify[ing], based on the interpretation of the first utterance, content to dedicate to a recipient,” “initiat[ing], based on the identified content, a dedication to the recipient,” “receiv[ing] a second utterance to be associated with the dedication,”

“determin[ing], based on a processing of the second utterance by the speech recognition engine, one or more words or phrases of the second utterance,” “provid[ing] the one or more words or phrases of the second utterance as textual annotations within metadata of the content,” and “send[ing] information to enable the recipient to access the content and the second utterance.” *Id.*, 37:25-55 (claim 1).

25. In my opinion, a POSITA at the priority date of the '025 patent would have had a Bachelor's degree in computer science, computer engineering, electrical engineering, or a 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. Additional experience can substitute for the level of education, and vice-versa.

26. I had at least the requisite level of skill with respect to the subject matter of the challenged patent at the time of the priority date as demonstrated by my qualifications listed above and my CV (EX1023).

D. Claim Construction

27. I understand that in *inter partes* review (“IPR”) proceedings, claims are construed using the same claim construction standard that district courts use, which I understand is called the *Phillips* claim construction standard, including construing the claim in accordance with the plain and ordinary meaning of such claim as

understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.

28. I applied plain-and-ordinary meaning to all terms of the challenged claims for my analysis in this proceeding.

29. I understand that a district court in *VB Assets, LLC, v. Amazon.com, Inc., et al.*, C.A. No. 19-1410 (MN) (D. Del.) construed the term “speech recognition engine” as “software or hardware that recognizes the words or phrases in the natural language utterance” for U.S. Patent Nos. 7,818,176, 8,886,536, and 9,269,097, which share an inventor and disclosure with the challenged patent. EX1021, 2. My opinion is that this construction is consistent with the plain and ordinary meaning of this term.

30. I am not an attorney, and my opinions are limited to what I believe a POSITA would have understood the meaning of certain claim terms to be, based on the patent specification and prosecution history in light of the POSITA’s knowledge.

V. OVERVIEW OF THE ’025 PATENT

31. The ’025 patent was filed as Application No. 12/943,699 (the “’699 application”) on November 10, 2010. The ’699 application claims the benefit of U.S. Provisional Application No. 61/259,820, filed on November 10, 2009.

32. The ’025 patent is directed to a system and method for providing “a natural language content dedication service” that receives and processes “requests

to dedicate content” and delivers “customized content to the recipients of the dedications.” EX1001, Abstract, 3:14-28. A natural language utterance comprising the request to dedicate content is interpreted using well-known interactive speech technology.² An Automatic Speech Recognizer (ASR) “receiv[es] the electronic signal corresponding to the utterance” from a voice-enabled device of a user and “generate[s] one or more preliminary interpretations of the utterance.” *Id.*, 8:17-19. The ASR provides the interpretation(s) to a conversational language processor, which “analyze[s] the preliminary interpretations of the utterance and any accompanying non-voice interactions to determine the intent” of the utterance. *Id.*, 11:16-18. For example, for the utterance “Find ‘Superstylin’ by Groove Armada,” the conversational language processor “may determine that the word ‘Find’ indicates that the most likely intent of the interaction includes a search, while the word ‘Superstylin’ and the phrase ‘Groove Armada’ provide criteria for the search.” *Id.*, 28:24-34.

² As I discuss in more detail below, interactive speech technology was well known by 2009. *See, e.g.*, EX1012, 879-982 (Chapters 17-18); EX1008, [0006] (“Much work covering multiple methods has been done in the fields of natural language processing and speech recognition. Speech recognition has steadily improved in accuracy and today is successfully used in a wide range of applications.”).

33. The user further provides a command to dedicate the identified content to a recipient, such as speaking by the utterance “Share this with my wife” or “Dedicate it to Charlene.” *Id.*, 28:55-57. A content dedication application prompts the user to provide a natural language dedication utterance (e.g., “Dear Charlene, I was listening to this song and I thought of you. Enjoy!”) and “route[s] the request, including the dedication utterance, [] additional tags (if any), and [] information identifying the recipient” to a content dedication system to “process a transaction for the content dedication request.” *Id.*, 29:16-39, 34:11-12. The content dedication system “associat[es] the dedication utterance” and any additional tags with the dedicated content and “send[s] a content dedication message to the recipient of the dedication,” e.g., by providing a “link that the recipient can select to stream, download, or otherwise access the dedicated content, the dedication utterance, etc.” *Id.*, 36:18-19, 36:33-39, 36:44-46.

VI. PROSECUTION HISTORY

34. I reviewed the prosecution history of the '025 patent. EX1003. The Applicant filed an information disclosure statement (“IDS”) with the '699 application citing several hundred references. *Id.*, 774-785. This IDS listed the Kennewick reference (EX1008) asserted in this petition. It also listed many other references naming Kennewick as an inventor. *Id.*, 782-85. The Applicant

subsequently filed additional IDS's listing hundreds more references. *Id.*, 451, 324-334, 244-254, 171, 102.

35. In an Office Action dated October 1, 2012, the Examiner rejected claims of the '699 application under 35 U.S.C. § 101 as directed to ineligible subject matter and under 35 U.S.C. § 112 as indefinite. *Id.*, 388-404. The Examiner also rejected the claims under 35 U.S.C. § 102 and § 103 over Kennewick (EX1008) alone or in various combinations with Gold (EX1005), Mahajan (U.S. Publication No. 2008/0140385), Caine (U.S. Publication No. 2010/0036967), and Ginter (EX1009). EX1003, 395-402. The Examiner cited Kennewick alone or in combination with Gold against the independent claims. EX1003, 395-400.

36. The Applicant responded on January 30, 2013 with amendments to the claims and stated that the amendments overcame the rejections under § 101 and § 112. *Id.*, 306-19. The Applicant introduced the "identify" and "initiate" limitations, versions of which ultimately issued as limitations 1.b.(v) and 1.b.(vi) of the '025 patent, and argued that the cited combination of Kennewick and Gold did not disclose these features. *Id.*, 307, 316. The Applicant further argued that dependent claim 6 was separately allowable because Mahajan did not teach annotating the content with the second utterance. *Id.*, 317.

37. In a Final Office Action issued March 5, 2013, the Examiner maintained the rejections under § 103. *Id.*, 167-98. Brachman (U.S. Patent No.

6,704,576) was added to the independent claim § 103 rejection for the newly added “identify” limitation, and the Examiner cited Gold as disclosing the “initiate” limitation. *Id.*, 191-92. The Examiner also added Anderson (U.S. Patent No. 8,521,315) to the § 103 rejections of dependent claims 21 and 22, which issued as claims 15 and 16. *Id.*, 196.

38. The Applicant filed a Notice of Appeal on June 5, 2013. *Id.*, 176. In its Appeal Brief, the Applicant argued that the cited combination of Kennewick-Gold-Brachman did not teach the “identify” step of claim 1. Namely, the Applicant argued that “while the natural language search in Brachman may produce content results, the natural language search does not identify either (i) the content to dedicate or (ii) the recipient to which the content is to be dedicated.” *Id.*, 151-52 (emphasis in original). The Applicant also separately addressed the rejections of several dependent claims, arguing with respect to claim 6 that Mahajan’s “[p]resentation of a transcription of content on a user interface as the content is being played” is not “annotation of content...with an utterance..., much less annotation of content with a second utterance that is to be associated with a dedication to a recipient.” *Id.*, 152-53 (emphasis omitted).

39. In his Answer, the Examiner cited two separate rationales supporting the rejection of claim 1. First, the Examiner asserted that the combination of Kennewick and Gold read on the “identifying” step in view of Kennewick’s natural

language processing and Gold's disclosure that a user can select a song to dedicate to a recipient. The Examiner explained that "using Kennewick to process a natural language utterance, and route the utterance as input to the system of Gold, would have been recognized by those of ordinary skill in the art as resulting in an improved system that would allow selection of content based on natural language." *Id.*, 133-35.

40. Second, the Examiner argued that "Brachman teaches: a user may request data using a natural language search and a voice means may be used to enter the user's natural language request" and asserted that "[s]ince each individual element and its function are shown in the prior art,...the difference between the claimed subject matter and the prior art rests not on any individual element or function but in the very combination itself that is in the substitution of content dedicated to the recipient for broadly defined content." *Id.*, 136.

41. After the Applicant filed a Reply Brief on February 27, 2014, the Patent Trial and Appeal Board (the "Board") conducted an oral hearing on April 5, 2016. *Id.*, 63-73, 172-82, 207-17. The Applicant asserted in its Reply Brief and at the hearing that the independent claims required that "both the content to be dedicated and the recipient are identified based on a natural language processing of one or more words or phrases of a first utterance," which the Applicant argued the cited references did not teach. *Id.*, 116, 72 (emphasis in original).

42. The Board disagreed, affirming-in-part the Examiner's rejections. *Id.*, 74-88. Specifically, with respect to the Kennewick-Gold combination, the Board noted that

Appellants' argument is not persuasive at least [because] it is not commensurate with the language of the claim. Appellants assert that claim 1 requires that both the content to be dedicated and the recipient are identified based on natural language processing of one or more words or phrases of the first utterance (Reply Br. 3-4). But claim 1 merely recites that the one or more words or phrases of the first utterance are processed to identify "content to dedicate to a recipient," i.e., that the first utterance is processed to identify the content, e.g., the particular song, that will be the subject of the dedication.

Id., 81. The Board also rejected the Applicant's arguments that Brachman's search produces a set of content results and not a particular content item of interest: "We also fail to see why, and Appellants do not explain why, Brachman's natural language search query could not be used to identify a single content item, rather than a set of content results, which single content item is the particular content item of interest." *Id.*, 82.

43. Nor did the Board find persuasive the Applicant's arguments for claim 6, agreeing with the Examiner that Gold disclosed this limitation. *Id.*, 84. The Board reversed the rejections of dependent claims 21-23. *Id.*, 85-88.

44. Following the Board’s decision, the Applicant and the Examiner conducted interviews on June 22 and June 30, 2016. *Id.*, 59-61, 48. The Examiner subsequently issued a Notice of Allowance on July 22, 2016. *Id.*, 33-47. The Notice of Allowance contained an Examiner’s Amendment that incorporated into the independent claims the limitation “provid[ing] the one or more words or phrases of the second utterance as textual annotations within metadata of the content,” previously recited in claim 23, as well as additional limitations describing the steps of the natural language processing. *Id.*, 38-44.

VII. RELATED LITIGATION

45. I understand that the ’025 patent is the subject of pending litigation involving Petitioner and Patent Owner: *VB Assets, LLC v. Amazon.com Services LLC*, No. 1:24-cv-00839-MN (D. Del., July 18, 2024) (the “parallel litigation”). I understand that Patent Owner sued Petitioner for alleged infringement of the ’025 patent. *See* EX1022.

46. I also understand that the ’025 patent is currently involved in another litigation: *VB Assets LLC v. SoundHound AI Inc.*, Case No. 1:24-cv-01279 (D. Del.).

VIII. PRIOR ART TO THE CHALLENGED PATENT

A. Technical Background

47. In this section, I discuss the state of the art relevant to the subject matter of the ’025 patent. The ’025 patent describes ideas related to (i) a human user interacting with a system via speech recognition, (ii) the system analyzing the

recognized words of the user's speech to identify content for dedication, (iii) the system processing a transaction for the identified content, (iv) the system analyzing recognized words from a second user speech input to annotate the content for dedication, and (v) the system delivering the annotated content to a recipient of the dedication. These ideas were well known in the prior art at the time of the '025 patent.

48. In interactive conversational systems involving a human user and a system, a dialog (or conversation) can be represented as a series of utterances, where each utterance is alternately generated by the human and by the system. A simple example of a dialog might be the following:

User: "I would like to fly to New York."

System: "When would you like to travel?"

User: "This Friday please."

System: "What is your departure airport?"

And so on. In this example there are two utterances by the user and two by the system. Technologies and systems to broadly support such interactive conversations had been under development and were well known prior to 2009, both in academic research laboratories and in industry. *See, e.g.,* EX1010; EX1011; EX1012, 879-982 (Chapters 17-18).

49. These types of interactive conversational systems, as developed and implemented by 2009, relied on a number of related technologies that I will briefly describe below. The primary technologies are automated speech recognition (“ASR”), natural language understanding (“NLU”), dialog management (“DM”), and user modeling (“UM”). Broadly speaking, ASR involves recognizing words from spoken audio for each user utterance, NLU is the process of determining the meaning of the recognized words in the utterance, DM involves generating appropriate response utterances by the system and managing the conversation with the user across multiple utterances, and UM is the process by which the system keeps track of information about a user, e.g., what their typical goals and preferences are.

50. These terms are not the only ones used to describe these technologies. Other terms were (and are still) used by researchers and practitioners in this general context. One such term is “Spoken Language Understanding” (“SLU”) which typically encompasses aspects of ASR, NLU, and DM. Nonetheless, the terms ASR, NLU, DM, and UM were in broad use in 2009 and would have been well known to a POSITA at that time. Thus, I focus on these terms in my discussion since these terms are generally useful for describing the technical background for the challenged patent.

51. In 2009, ASR involved a sequence of steps consisting of (i) recording and digitizing speech waveforms spoken by a user, (ii) converting the digitized

speech waveforms into numerical features that (typically) represent the frequency content of very short segments of the speech waveform, (iii) automatically classifying these features into corresponding phonemes (basic units of sounds within words), and (iv) recognizing specific words based on sequences of phonemes. This process has been the subject of extensive research and commercial activity since the 1960s. For example, the well-known text “SLP” devotes Chapters 5 through 13 (approximately 480 pages) just to ASR. EX1012, 227-711.

52. The output of the ASR component of a conversational interface system consists of the ASR’s best estimates of what words a user spoke in an utterance. The recognized words alone are typically not sufficient for an algorithm to understand the user’s goals and intent. Human language is very flexible, and there are many ways to express the same meaning or intent. The technology of NLU consists of a broad range of algorithms and methods that allow a computer system to interpret the meaning of the recognized words from the ASR. *Id.*, 893-912. For example, consider the simple phrase “I would like to fly to New York” from the dialog example described earlier. And assume that the system is a travel reservation system that can find flights, make flight reservations, pay for the flights, suggest or advertise additional travel services, and so on. Such a system would have an internal representation (e.g., a database) of items such as names of airports, locations of airports, lists of flights, actions the system can take (e.g., book flights), and so on.

Thus, it would need to interpret the user's words "like to fly" as an implicit request for the system to find a flight. It would also need to interpret the word "I" to mean that the flight is for the individual who is speaking. And it would need to understand that "New York" is the name of a location, and that it is not the name of a specific airport but is a city with multiple airports associated with it. There are different approaches associated with developing NLU technologies that can robustly handle such situations, and by 2009 multiple such systems had been developed that demonstrated the necessary NLU capabilities for interpreting the meaning of spoken words in conversational dialog. *See, e.g., id.*, 945-982 (Chapter 18); EX1011; EX1013, 18:39–19:37.

53. As I describe above, the ASR and NLU components of a conversational interface system focus on recognizing words and interpreting these words on a per-utterance basis. DM (Dialog Management) technology addresses the problem of tracking different aspects of a dialog across multiple utterances. For example, for a travel reservation system interacting with a user via speech, the DM component of the system gradually accumulates the relevant pieces of information from the user (dates, times, arrival/departure airports, seating preferences, confirmation of booking, purchase information) over the course of a dialog, i.e., over a sequence of utterances generated the user. Thus, DM technologies involve creating internal representations of concepts such as goals for a user and methods to satisfy those

goals. One such method includes directly asking the user for specific information, which in turn involves automatically generating the appropriate words for the system's response (utterance) to the user, given what the system knows already (e.g., "What time would you like to leave for your flight to New York?") and then converting that into a spoken response that the user can hear. Another type of DM method involves inferring the relevant information from knowledge sources available to the system. For example, the DM might assume from information in the user's profile that the departure airport is likely to be the one closest to where the user lives. And the DM could infer the likely destination airport if non-stop flights from the user's home airport only go to a specific airport in the New York area. DM technology involves various software components that can make decisions about what type of information to request from the user, how to present this information to the user, in what sequence to request this information, when this information might be inferred from other sources, and so on. DM technology that exhibited these characteristics was well known by 2009 having been developed in research laboratories and in commercial systems before that time. *See, e.g.,* EX1012, 912-927 (Chapter 17.6-17.7), 945-982 (Chapter 18); EX1011.

54. By 2009, systems and methods for a conversational voice user interfaces were well known. Indeed, as Kennewick recognizes:

Much work covering multiple methods has been done in the fields of natural language processing and speech recognition. Speech recognition has steadily improved in accuracy and today is successfully used in a wide range of applications. Natural language processing has previously been applied to the parsing of speech queries.

EX1008, [0006]. Speech recognition software was also commercially available.

55. In addition, by 2009, Internet-based media playback platforms and services were in wide use and had become a popular alternative or supplement to traditional physical distribution of music. The “combination of the MP3 format with the Internet” in the late 1990s “revolutioniz[ed] the music world” by “enabling powerful online distribution of music.” EX1014, 1. The ability of the MP3 format to achieve “near-CD quality at 128 Kbps” provided a “favorable trade-off of an insignificant loss in quality for a large degree of compression, not to speak of the growing capacity and decreasing prices of hard drives,” and allowed for “online-storage of CD-quality music.” *Id.*

56. While peer-to-peer music sharing using platforms like Napster was a popular early approach to digital music distribution, Apple made significant traction in the field with the launch of the iPod MP3 player in 2001 and the iTunes music store in 2003, popularizing digital (and legal) downloads from a centralized server. EX1015, EX1016; EX1017. Other platforms in the paid digital music space, like

Rhapsody and Yahoo! Music Unlimited, enabled subscription-based or ad-supported on-demand streaming of songs, allowing a user to query the system for a specific track or artist and stream identified search results in a web browser or standalone application. EX1018, [0003].

57. Some of these services allowed users to share and exchange music and other media items. For example, Svendsen recognized in 2008 that “[d]ue to the large amount of the accessible digital media, recommendation technologies are emerging as an important enabler to assist users in identifying and navigating large databases of available media.” *Id.*, [0003]. Svendsen describes a form of media item recommendation that includes a media item identifier (i.e., to allow the recipient “to access the media item for playing”) as well as a media item annotation from the recommender. *Id.*, [0033]-[0034]; *see also id.*, [0053]. The “exchange of media item recommendations among the users,” particularly among users having an existing relationship “help[s] users identify and select media items of interest for usage and/or play.” *Id.*, [0004], [0003]. Svendsen also explains that allowing a user to “include personal information and annotations in the media item recommendation” would “develop and strengthen the user interaction, which would promote the continued use of the [platform] for media item recommendations and the acquisition and playing of the media items recommended.” *Id.*, [0005].

58. Apple also integrated music sharing functionality into iTunes by 2005, “enabling users to purchase tracks and albums identified in” a playlist generated and published by another user. EX1019, [0013], [0010]. Gold, published in 2002, described a content dedication service whereby a sending user could “send dedications of popular songs through a communications medium such as the Internet,” recognizing that “song dedications can be an effective way to convey and share complex emotions and memories when it might not be possible, or desired, for the sender and recipient to communicate directly.” EX1005, Abstract, [0006]. And Wang, published in 2004, describes an early version of the Shazam music identification service, which identified songs from short samples of audio and enabled users to purchase and share an identified track. EX1007, [0012], [0043]. A user could, for instance, “instruct the service to send a recommendation for the identified song to a friend via email or SMS or a clip of the song to the friend’s mobile telephone.” *Id.*, [0043].

59. By 2009, it was also well-understood that the types of conversational systems that I described above, involving combinations of ASR, NLU, DM, and UM, could be directly applied in the context of digital music distribution and sharing. For example, Wang discloses that its music identification service “[e]mploy[ed] an interactive voice response unit (‘IVR’)” that enabled users to “immediately purchase a recording of [an] identified song,” using “voice recognition

methodologies...to translate the user's voice responses into usable system inputs.” *Id.*, [0012], [0084]. And Stefik, published in 2009, discloses a music recommendation system that “provide[s] voice recognition by which the recommender could speak out the music selection and recipient” and “optionally annotate a message to the recommendation...either in text or voice.” EX1020, [0048]-[0049].

B. Summary of the Prior Art

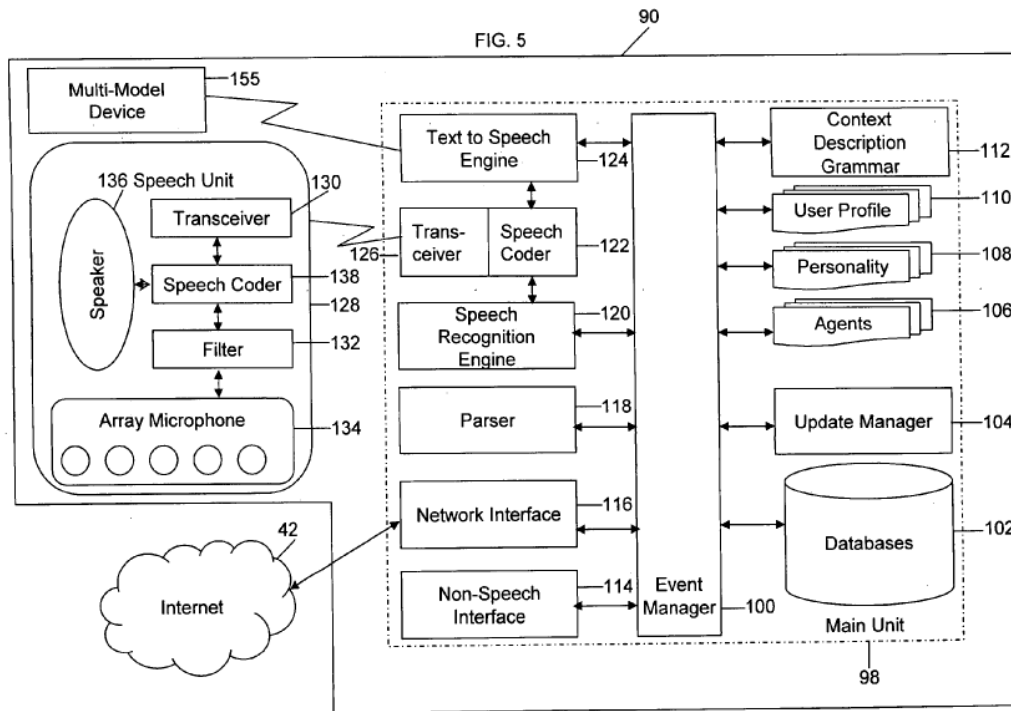
1. Weider (EX1004)

60. Weider, U.S. Publication No. 2007/0050191, is entitled “Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions.” Weider was filed August 29, 2005 and published March 1, 2007. EX1004. I understand that both Weider and the patent application that issued as the ’025 patent were filed by VoiceBox Technologies, Inc. (“VoiceBox”). Additionally, I understand that the ’025 patent cites to Weider and purports to incorporate Weider by reference. EX1001, 29:56-64. I also understand Weider is prior art because it was published more than a year before the November 10, 2009 priority date of the challenged patent. I form no opinion as to whether Weider is entitled to an earlier priority date.

61. Weider discloses a “speech interface” that “enables mobile users to submit natural language speech and/or non-speech questions or commands in a wide

range of domains.” EX1004, [0001]. The speech interface “identifies and uses context...and user specific profile data” to interpret a user’s request and provide responses. *Id.*, Abstract.

62. As FIG. 5 of Weider illustrates below, a speech recognition engine 120 “recognizes words and phrases” in a user’s natural language utterance. *Id.*, [0187]. Then, parser 118 determines a context for the user’s utterance based on the recognized words and phrases of the utterance by “applying prior probabilities or fuzzy possibilities to keyword matching.” *Id.*, [0192]. The parser 118 may “use[] a scoring system to determine the most likely context or domain for a user’s question and/or command” by “weighing a number of factors including, the user profile 110, the domain agent’s data content and previous context.” *Id.*, [0193]. For example, in one dialog, “the keyword[] ‘temperature’ implies a context value of weather for the question” but in “a different dialog, the keyword ‘temperature’ can imply a context for a measurement” such as body temperature. *Id.*, [0192]. “Once the context for the question or command has been determined, the parser 118 can invoke the correct agent 156, 150.” *Id.*, [0195].



Id., Fig. 5.

63. Weider's speech interface may be used in a variety of different applications, including to control "multimedia entertainment electronics," "communications devices such as cell phones, voice mail systems, ... text or instant messaging systems, ... email systems," etc. *Id.*, [0044]-[0045]. For example, the user may use the speech interface to "play a specific song title." *Id.*, [0139].

2. Gold (EX1005)

64. Gold, U.S. Publication No. 2002/0032752, is entitled "Method and System for Electronic Song Dedication." Gold was filed June 7, 2001 and published March 14, 2002. Gold claims the benefit of Provisional Application No. 60/210,668 filed on June 9, 2000. EX1005. I understand Gold is prior art because it was

published more than a year before the November 10, 2009 priority date of the challenged patent. I form no opinion as to whether Gold is entitled to an earlier priority date.

65. Gold discloses “[a] system and method to send dedications of popular songs through a communications medium such as the Internet.” *Id.*, Abstract. A sender navigates to a “musical jukebox Web site” and “enters song search criteria 14 such as a song title, a performing artist’s name, or the year a song was first released.” *Id.* Abstract, [0031]. The system provides a list of songs matching the search criteria, and “[t]he sender selects one of the songs from the received list to be the dedicated song 18.” *Id.*, [0031]; *see also id.*, Figs. 4-6. After the user enters “transaction information including the email address of the sender and the recipient,” the jukebox Web server “sends the sender’s and the recipient’s email addresses and a Uniform Resource Locator (URL) pointing to the dedicated song to [a] dedication Web server,” which returns a Personal Identification Number (PIN) “given as a unique identifier to the sender.” *Id.*, [0031], [0037].

66. The jukebox Web server “sends the PIN along with instructions for recording a voice message to the sender Web client.” *Id.*, [0038]. The sender “use[s] an audio device such as a telephone to access [a] voice message CT application and make a voice message.” *Id.* Alternatively, “if the sender’s computer has the appropriate hardware and software, a voice message can be recorded using the

sender's computer." *Id.*, [0007]; *see also id.*, [0070]. The dedication message, "when combined with the dedicated song will create a complete dedication." *Id.*, [0031]. "The PIN is used to make sure that the appropriate sender and voice message correspond to one another." *Id.*, [0038].

67. The dedication Web server associates the voice message with the song and generates and sends to the recipient email address a "notification inform[ing] the recipient that a dedication has been made by the sender to the recipient." *Id.*, [0042]; *see also id.*, Fig. 11. The email notification "contain[s] the URL of a play dedication server script located on the dedication Web server," which "uses the PIN to determine the URL of the voice message served by the dedication Web server and the URL of the dedicated song served by the jukebox Web server." *Id.*, [0043]. The play dedication server script "returns a Web page and a dedication client script," which, when executed, causes "the voice message and the dedicated song [to be played] as a single dedication by the recipient's Web client." *Id.*

3. Martinez (EX1006)

68. Martinez, U.S. Publication No. 2007/0083762, is entitled "Set of Metadata for Association with a Composite Media Item and Tool for Creating Such Set of Metadata." Martinez was filed October 10, 2005 and published April 12, 2007. EX1006. I understand Martinez is prior art because it was published more

than a year before the November 10, 2009 priority date of the challenged patent. I form no opinion as to whether Martinez is entitled to an earlier priority date.

69. Martinez provides a “tool for creating a composite media item with an associated set of metadata.” *Id.*, [0006]. The tool includes “a component for extracting a portion of a first media item having first metadata and for extracting a portion of a second media item having second metadata.” *Id.*, [0018]. Martinez combines the first and second portions to form a composite media item (e.g., a “remix” of the first and second media items). *Id.*, [0104]; *see also id.* [0103] (“[T]he composite media item 200, is based on two media item portions; an a cappella track from Gwen Stefani 204 and a backing track from the Gorillaz 205.”).

70. The tool further includes a component for analyzing the metadata associated with the first and second media items and “extract[ing] portions of the first and second metadata to form a new set of metadata for association with the composite media item.” *Id.*, [0018]. The extracted portions are “inserted[...]into at least one data container.” *Id.*, [0020]. The data container includes both content (e.g., audio or images associated with the first and second media items) and associated metadata. *Id.*, [0035], [0045]. Martinez further discloses that the containers “may be created using a variety of methods such as XML, using linear strings of data with internal header information, or software objects such as those utilized in the Java or C++ programming language.” *Id.*, [0044]; *see also id.*, [0004] (“[N]ew digital media

models, such as the MPEG-21 standard or enhanced ID3 tags, enable the binding of detailed descriptions or metadata about an item of digital media.”). The data container includes “informative tags associated with the composite media item.” *Id.*, [0016]. These “tags” store the metadata that is associated with the media items. *Id.*, [0004] (“[M]etadata is typically entered...into whatever tags are available, such as, for example, ID3 tags in MP3 files.”).

4. Wang (EX1007)

71. Wang, U.S. Publication No. 2004/0199387, is entitled “Method and System for Purchasing Pre-Recorded Music.” Wang was filed April 26, 2004 and published October 7, 2004. EX1007. I understand Wang is prior art because it was published more than a year before the November 10, 2009 priority date of the challenged patent. I form no opinion as to whether Wang is entitled to an earlier priority date.

72. Wang discloses a “signal identification service” that enables a user to identify and purchase a song through an “interactive voice response unit (‘IVR’).” *Id.*, [0012]. A user of the service “captures a relatively short sample of music being played over a music playing device,” “plac[es] a call with a mobile telephone” and “play[s] the music into the handset of the telephone.” *Id.*, [0012]. “A system at the other end of the telephone identifies the song to the user in real-time” by comparing

the received audio sample “to those music recordings in its database and then identifies the song to the user.” *Id.*, [0012], [0024].

73. Once the service identifies the song, “the user is then provided the opportunity to immediately place an order for a recording of the identified music,” such as “an album on CD or tape that contains the selected track” or “digital download[] to users’ PCs, wireless internet devices, hand held personal computers, palm pilots, or mobile telephones.” *Id.*, [0024], [0025], [0026]. “The user’s account is then debited” based on the selected purchase option “and the selection is forwarded to the user.” *Id.*, [0027]. The service therefore “enabl[es] music retailers to take advantage of impulse purchase phenomena” by “providing a mechanism that guarantees that the user is at the height of interest in the particular music” when the purchase options are provided. *Id.*, [0010], [0012].

74. Wang’s signal identification service is further configured to “implement tasks and instructions directed by the user,” including by “allow[ing] a user to interact with other[] users, non-users, or potential users of the service.” *Id.*, [0042]. For example, “the user may instruct the service to send a recommendation for the identified song to a friend via email or SMS or a clip of the song to the friend’s mobile telephone.” *Id.*, [0043]. Users may also “create their own music compilations and share them with others, using for example, MP3 storage and access protocols.” *Id.*, [0041].

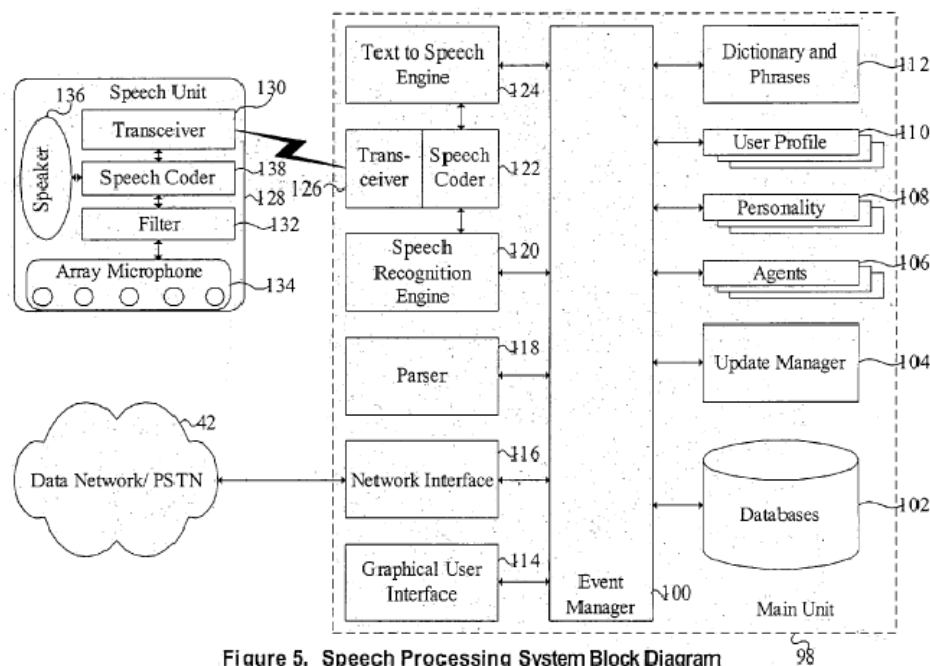
5. Kennewick (EX1008)

75. Kennewick, U.S. Publication No. 2004/0193420, is entitled “Mobile Systems and Methods for Responding to Natural Language Speech Utterance.” Kennewick was filed July 15, 2003 and published September 30, 2004. Kennewick claims the benefit of Provisional Application No. 60/395,615 filed July 15, 2002. EX1008. I understand Kennewick is prior art because it was published more than a year before the November 10, 2009 priority date of the challenged patent. I form no opinion as to whether Kennewick is entitled to an earlier priority date.

76. Kennewick discloses a “speech interface” that enables “mobile users to ask natural language speech questions or give natural language commands in a wide range of domains.” *Id.*, [0002]. The speech interface makes “significant use of context...and user specific profile data” to interpret a user’s request and provide responses. *Id.*, Abstract.

77. As FIG. 5 of Kennewick illustrates below, a speech recognition engine 120 “recognize[s] words and phrases” in a user’s natural language utterance. *Id.*, [0155]. Then parser 118 determines a context for the user’s utterance based on the recognized words and phrases by “applying prior probabilities or fuzzy possibilities to keyword matching.” *Id.*, [0160]. “The parser 118 may use a scoring system to determine the most[] likely context or domain [by]...weighting a number of factors including, the user profile 110, the domain agent’s 156 knowledge[,] and previous

context.” *Id.*, [0160]-[0161]. For example, in one dialog “the keyword[] ‘temperature’ implies a context value of weather for the question” but in “a different dialog, the keyword ‘temperature’ can imply a context for a measurement” such as body temperature. *Id.*, [0160]. “Once the context for the question or command has been determined, the parser 118 can invoke the correct agent 156, 150.” *Id.*, [0162].



Id., Fig. 5.

78. Additionally, Kennewick discloses that the speech interface may be used in a variety of different applications, including to control “multimedia entertainment electronics,” “communication devices” such as “cell phones, voice mail, ...text or instant messaging systems, ...email systems,” etc. *Id.*, [0043],

[0044]. For example, the user may command the system to play “a particular piece of music.” *Id.*, [0018].

6. Ginter (EX1009)

79. Ginter, U.S. Patent No. 5,892,900, is entitled “Systems and Methods for Secure Transaction Management and Electronic Rights Protection.” Ginter was filed August 30, 1996 and issued April 6, 1999. EX1009. I understand Ginter is prior art because it was published more than a year before the November 10, 2009 priority date of the challenged patent. I form no opinion as to whether Ginter is entitled to an earlier priority date.

80. Ginter describes a “virtual distribution environment (VDE) that may enforce a secure chain of handling and control, for example, to control and/or meter or otherwise monitor use of electronically stored or disseminated information.” *Id.*, Abstract. VDE content control information may have “differing metering, budgeting, billing and/or payment models,” such as content usage payments made “through a clearinghouse, or directly, to different content providers for different portions.” *Id.*, 33:31-35.

81. Ginter discloses that the VDE “can protect the rights of parties who have...commercial interests in electronically distributed information,” by “ensur[ing], for example, that parties, will be paid for use of distributed information in a manner consistent with their agreement.” *Id.*, 4:28-34. For example, “an

electronic agreement between a content creator and a distributor may establish both the price to the distributor for a creator’s content...and the number of copies of this object that this distributor may distribute to end-users.” *Id.*, 47:24-30. Ginter further discloses that an agreement may involve an end-user. For example, Ginter describes a “billing process” that “determines how much to charge for events” and “records and reports payment information.” *Id.*, 58:62-63; *see also id.*, 166:38-46.

IX. SUMMARY OF INVALIDITY GROUNDS

82. In my opinion, the challenged claims are invalid based on the following grounds:

- Ground 1: Weider in combination with Gold renders obvious claims 1, 3, 6-8, 10, 13-14, and 19;
- Ground 2: Weider in combination with Gold and Martinez renders obvious claims 2, 9, 17, and 18;
- Ground 3: Weider in combination with Gold and Ginter renders obvious claims 4 and 11;
- Ground 4: Weider in combination with Gold, Ginter, and Wang renders obvious claims 5 and 12;
- Ground 5: Weider in combination with Gold and Wang renders obvious claims 15-16;

- Ground 6: Kennewick in combination with Gold and Weider renders obvious claims 1, 3, 6-8, 10, 13-14, and 19;
- Ground 7: Kennewick in combination with Gold, Weider, and Martinez renders obvious claims 2, 9, 17, and 18;
- Ground 8: Kennewick in combination with Gold, Weider, and Ginter renders obvious claims 4 and 11;
- Ground 9: Kennewick in combination with Gold, Weider, Ginter, and Wang renders obvious claims 5 and 12; and
- Ground 10: Kennewick in combination with Gold, Weider, and Wang renders obvious claims 15-16.

X. INVALIDITY GROUNDS

A. Ground 1: Weider in combination with Gold renders obvious claims 1, 3, 6-8, 10, 13-14, and 19

1. Claim 1

[1.pre] “A system for providing a natural language content dedication service, comprising:”

83. To the extent that the preamble is limiting, Weider in combination with Gold renders it obvious. Weider discloses “[a] *system for providing a natural language...service.*” Weider describes an “interactive natural language system” that provides a “complete speech-based natural language command, algorithm and response environment.” EX1004, [0012], [0130]. Specifically, Weider’s system

“enables mobile users to submit natural language speech and/or non-speech questions or commands in a wide range of domains” and receive “a natural language speech response.” EX1004, [0001], [0014]. For example, “a user may utter ‘tune in my favorite radio station’” or “command the system to play a specific song title.” *Id.*, [0027], [0139].

84. For each utterance, Weider discloses that the system “processes the input” to interpret the user command, invokes an agent to obtain result data, and constructs a response to the user. *Id.*, [0014], [0028]; *see also* FIG. 5 (reproduced below, illustrating the interactive natural language processing system, which receives a user utterance via Speech Unit 128 and processes the input at Main Unit 98):

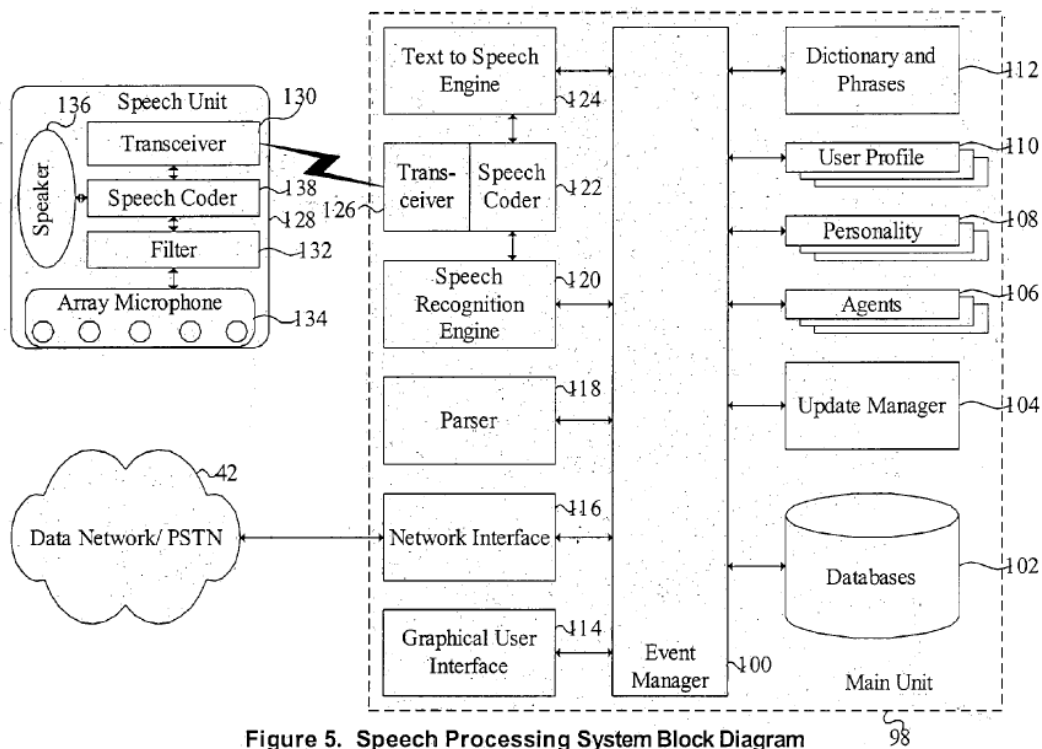


Figure 5. Speech Processing System Block Diagram

The response is passed to a text-to-speech engine, which “generates the utterances” that are “output by the speaker 136.” *Id.*, [0147]; *see also id.*, [0020], [0211].

85. Weider’s interactive natural language system is “[a] system for providing a natural language...service” because a user provides input to the system via a natural language utterance (e.g., a spoken command or query) and receives, from the system, a natural language response. *Id.*, [0014]. The system provides a “service” to users by processing and responding to user utterances. For example, when the user utterance is a command to “tune in my favorite radio station,” the system can process the request and tune to that station. *Id.*, [0027]. And when the

user “command[s] the system to play a specific song title,” the system can provide a song in response. *Id.*, [0139].

86. To the extent Weider does not explicitly disclose that the system “***provid[es] a ...content dedication service***,” the combination with Gold renders this obvious. Gold discloses a “song dedication system” by which a sender “create[s] a song dedication that can then be sent to a recipient.” EX1005, [0010], [0007]. The sender “uses the sender web client to access [a] jukebox Web server” containing “a database of songs for use in the dedication.” *Id.*, [0036], [0007]. Alternatively, “the sender, using a telephone calls ...[an] Interactive Voice Response (IVR) dedication system 32” and “enters a song search criteria...using the telephone’s keypad.” *Id.*, [0032]; *see also id.*, [0045], FIG. 1B. The database “is preferably searchable by criteria such as title, artist, and year the song was released.” *Id.*, [0007]. “Upon entering the search criteria, the sender receives a list of available songs meeting the criteria” and can select a song (content) for dedication, as shown in the following excerpt of FIG. 5 (annotations added³):

³ All annotations and emphases are added unless otherwise noted.

Jukebox: Search Results

Your search returned 1 matches.

410	ID:	228	405
415	Artist:	David Bowie	
420	Song:	Fame	
425	Year:	75	
	Audio:	Play Song	Dedicate Song

430

| [New Search](#) | [List All](#) |

400

Id. [0007]; *see also id.*, [0036], [0048].

87. Gold further discloses that the sender can record “a voice message to be appended to the dedicated song to create the desired dedication.” *Id.*, [0007]; *see also id.*, [0070] (“[T]he step of recording a voice message is completed by using a microphone and digitizer sound card connected to the sender host hosting the sender Web client.”). Once the sender creates the dedication, Gold’s dedication Web server “sends a first email recipient notification 245 to the recipient” to “inform[] the recipient that a dedication has been made by the sender to the recipient.” *Id.*, [0042]. The recipient can execute a dedication client script on their web client to play “the voice message and the dedicated song.” *Id.*, [0043].

88. A POSITA would have understood that Gold’s song dedication system is a “*content dedication service*” because it enables users (e.g., Gold’s sender) to

dedicate content (e.g., songs) to recipients. Indeed, Gold expressly refers to the content dedication functionality as a service. *Id.*, [0009] (“The dedication service provided by the invention may be used in any one of a number of different ways including various commercial transactions involving the Internet.”).

89. It would have been obvious to modify Weider to include Gold’s content dedication service and thereby provide a “*natural language content dedication service*.” Weider discloses a natural language service with music-related functionality, including functionality for playing specific song titles. EX1004, [0027] (“[A] user may utter ‘tune in my favorite radio station.’”), [0139] (“[T]he user may command the system to play a specific song title.”), [0119] (Weider’s “knowledge-enhanced speech recognition” identifies context categories associated with a command, including a “music” category). A POSITA would have been motivated to add Gold’s content dedication functionality to Weider’s service to enable users to share songs with others, as “song dedications can be an effective way to convey and share complex emotions and memories when it might not be possible, or desired, for the sender and recipient to communicate directly.” EX1005, [0006].

90. This is a combination of prior art elements according to known methods to yield predictable results. Weider’s system uses a flexible agent-based architecture to provide natural language services “in a wide range of domains.” EX1004, [0001] (“the invention enables mobile users to submit natural language speech and/or non-

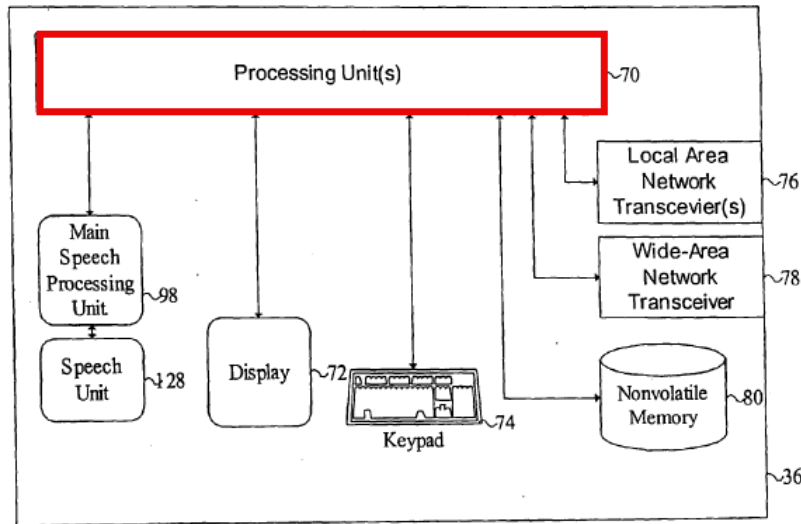
speech questions or commands in a wide range of domains”), [0150] (“generic and domain specific behavior and information may be organized into domain agents...the domain agents may include data that is needed to extend or modify the functionality of the system 90 in a current or new domain...[d]omain agents may access a plurality of sources that may provide various services”). Gold’s content dedication functionality—sending and dedicating a selected song to a recipient—is one type of domain. A POSITA would have found it obvious to use Gold’s functionality within Weider’s system, such as by using a domain agent like those taught by Weider, to provide the content dedication service taught by Gold. A POSITA would have found the result of the combination predictable because computer and Internet technologies were well understood and would have had a reasonable expectation of success for the same reason.

91. Alternatively, it would have been obvious to combine Gold and Weider for reasons like the Examiner identified for the Gold-Kennewick combination during prosecution of the ’025 patent. EX1003, 187-98 (Final Office Action), 129-44 (Examiner’s Answer). Specifically, it would have been obvious to a POSITA to modify Gold by substituting Weider’s natural language command for Gold’s UI-based instruction to send a dedication to a recipient. For example, it would have been obvious to use Weider’s natural language service to process a natural language utterance (e.g., a content dedication request) and to route that utterance as input, e.g.,

to identify content, to Gold's song dedication system. This is a simple substitution of one known element (Weider's method for receiving and processing natural language utterances) for another (Gold's song selection method). Moreover, as discussed above, Gold describes an embodiment in which the system is an "Interactive Voice Response (IVR) dedication system" with which the user interacts via telephone. EX1005, [0032]; *see also id.*, [0045] (The IVR features "allow[] the jukebox Web server to be accessed from a sender telephone."). As the Examiner explained in the context of the Kennewick-Gold combination, a POSITA would have recognized the combination of natural language processing with Gold's content dedication service as "resulting in an improved system that would allow selection of content based on natural language." EX1003, 133-35. Weider in combination with Gold therefore renders preamble [1.pre] obvious.

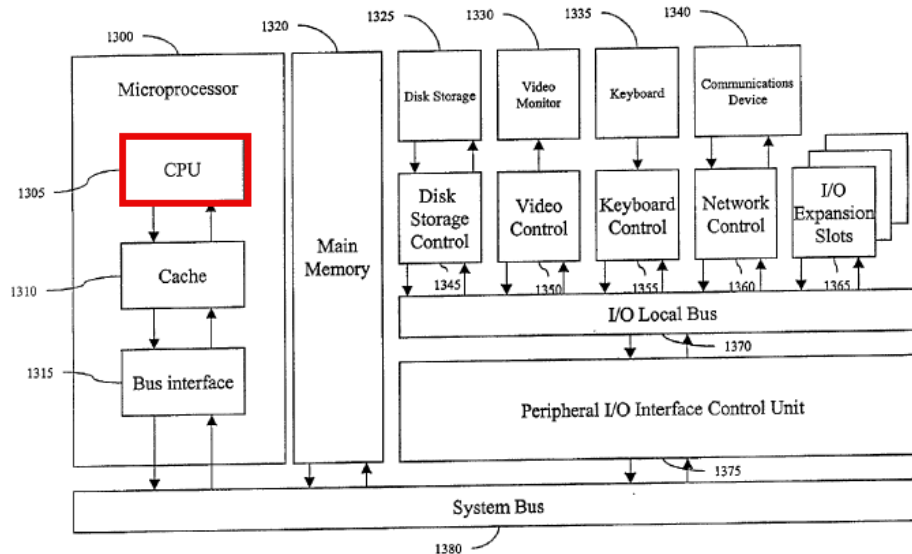
[1.a]: "one or more processors; and;"

92. Weider in combination with Gold renders this limitation obvious. Weider's interactive natural language system "may be incorporated into mobile devices or may be connected to the mobile device via a wired or wireless connection." EX1004, [0012]. Weider discloses that the mobile device, which is illustrated in FIG. 3, contains "one or more processing units 70," which "may include one or more central processing units." *Id.*, [0113]; *see also id.*, FIG. 3:



Likewise, in the embodiment that “uses a fixed computer 44,” Weider discloses a main speech processing unit that “interface[s] with one or more processing units 84,” which “may include one or more central processing units.” *Id.*, [0126]; *see also id.*, FIG. 4. A central processing unit is a processor. EX1024, 132 (defining a central processing unit as “the device that interprets and executes instructions” and a single-chip central processing unit as a “microprocessor”). Therefore, Weider renders obvious that the system comprises “*one or more processors.*”

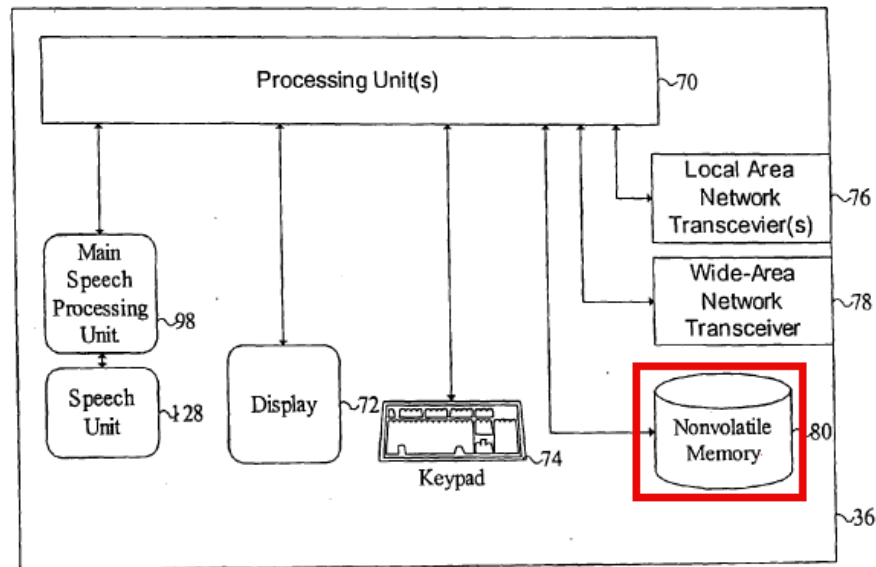
93. Gold also discloses “*one or more processors.*” Gold’s system includes a “general purpose computer suitable for use as a song dedication server.” EX1005, [0068]. The computer includes “a microprocessor 1300, comprised of a Central Processing Unit (CPU) 1310.” *Id.*; *see also id.*, FIG. 16:



The CPU is a processor. EX1005, [0068]; EX1024, 132. Weider in combination with Gold therefore renders limitation [1.a] obvious.

[1.b] “one or more non-transitory computer readable mediums storing executable instructions that when executed by the one or more processors cause the one or more processors to:”

94. Weider in combination with Gold renders this limitation obvious. Weider’s processing unit “may use one or more types of nonvolatile memory 80 for software and data storage. Suitable types of nonvolatile memory 80 may include flash memory and hard disk drives.” EX1004, [0113]; *see also id.*, FIG. 3:

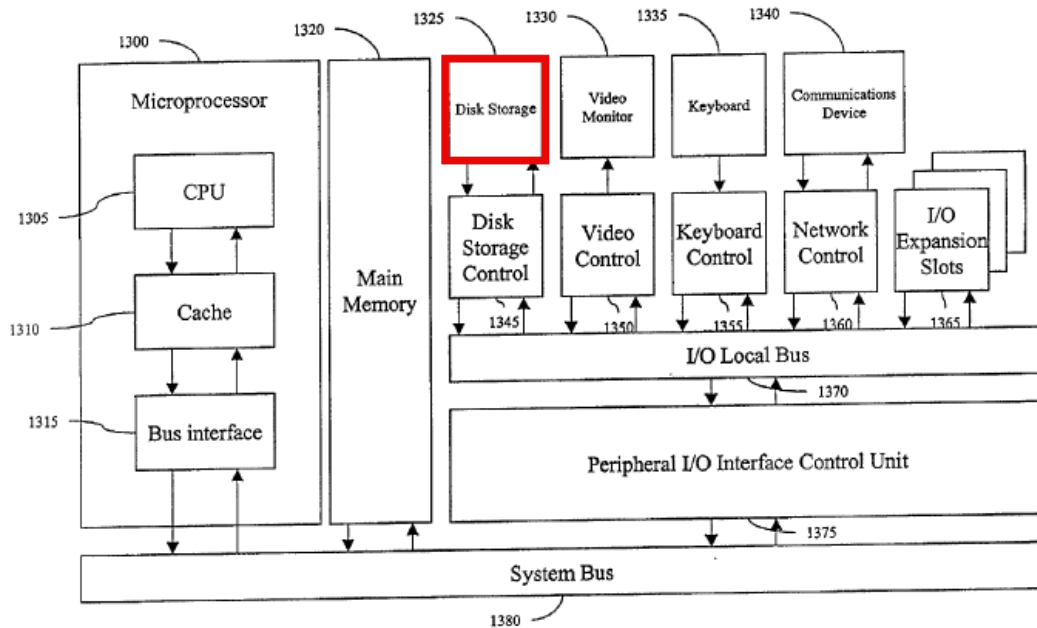


Likewise, in the embodiment that uses a “fixed computer 44,” Weider discloses “nonvolatile memory 94 for software and data storage.” *Id.*, [0126], FIG. 4. A POSITA would have understood that nonvolatile memory is a type of non-transitory computer readable medium because it stores instructions and data readable by the processor. Weider therefore discloses “*one or more non-transitory computer readable mediums.*”

95. Weider also discloses that the computer readable medium “*stor[es] executable instructions.*” For example, Weider’s speech processing unit contains various software modules, including a “speech recognition engine,” a “parser,” and a “text to speech engine.” *Id.*, [0013], [0020], [0167]. Weider also discloses that “domain specific behavior and information may be organized into data managers,” which are “autonomous executables that receive, process and respond to user questions, queries and commands.” *Id.*, [0023]; *see also id.* (“Data managers may

be complete packages of executable code.”). A POSITA would have understood that Weider’s software modules and data managers are executable instructions (i.e., instructions that are executed by the “*one or more processors*”) and are stored in the “*one or more non-transitory computer readable mediums*” (e.g., nonvolatile memory 80). A POSITA also would have recognized that the instructions, “*when executed by the one or more processors cause the one or more processors to*” perform programmed operations because this is how computers conventionally operate.

96. Gold also discloses “*one or more non-transitory computer readable mediums storing executable instructions that when executed by the one or more processors cause the one or more processors to*” perform programmed operations. Gold discloses a “disk storage device” that stores “[c]omputer program instructions 1397 implementing software objects.” EX1005, [0069]; *see also id.*, FIG. 16:



The instructions “are stored on the disk storage device until the microprocessor retrieves the computer program instructions and stores them in the main memory.”

Id. “The microprocessor then executes the computer program instructions stored in the main memory to instantiate a song dedication server.” *Id.*

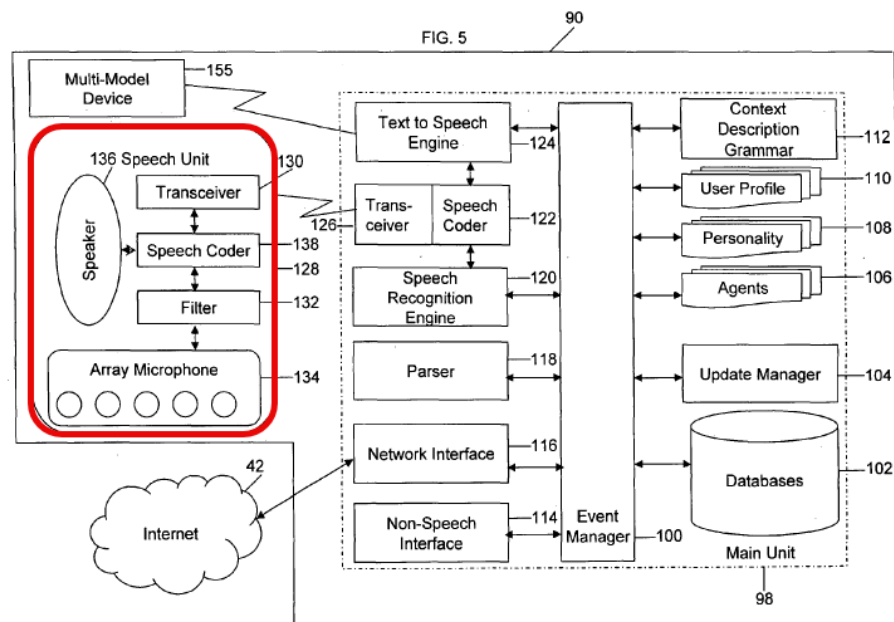
97. The disk storage device is a “*non-transitory computer readable medium*” and “*stor[es] executable instructions*” (“computer program instructions 1397”) that are “*executed by the one or more processors*” and “*cause the one or more processors to*” to carry out the operations specified by the instructions. Weider in combination with Gold therefore renders limitation [1.b] obvious.

[1.b(i)] “receive a first utterance that includes a natural language utterance;”

98. Weider in combination with Gold renders this limitation obvious. Weider’s interactive natural language system includes a “speech unit interface

device that receives spoken natural language requests, commands and/or other utterances from users.” EX1004, [0014]; *see also id.*, [0035] (“[T]he system may accept any natural language question or command.”). For example, a user may “command the system to play a specific song title” or “utter ‘tune in my favorite radio station.’” *Id.*, [0139], [0027]. A natural language command instructing the system to “play a specific song” is “*a first utterance that includes a natural language utterance.*”

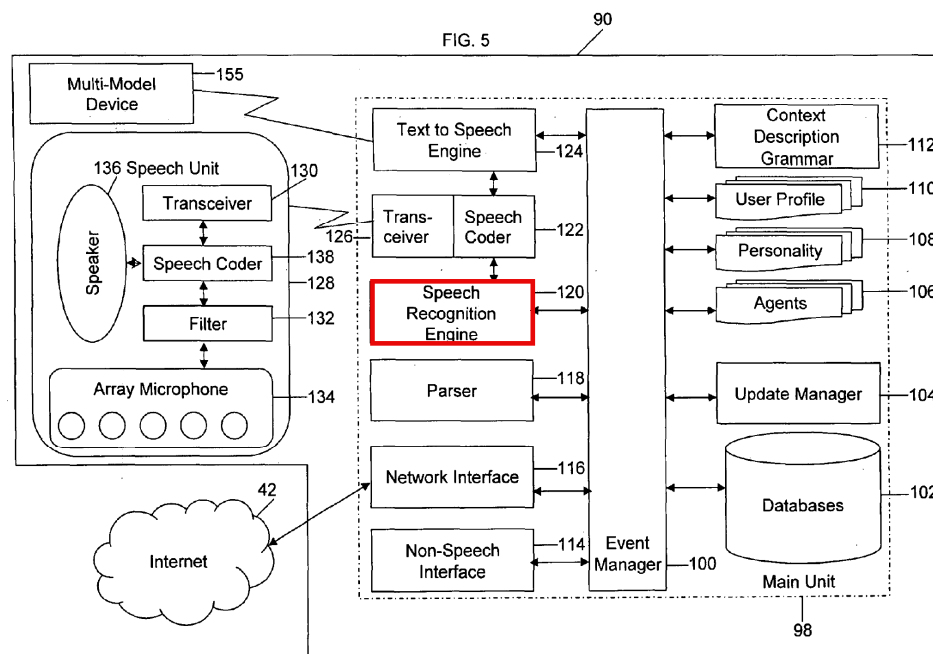
99. Weider also discloses “*receiv[ing]*” the “*first utterance.*” Weider’s speech unit interface device includes a speech unit 128, which includes “one or more microphones” to “receive the utterances from the user.” *Id.*, [0143]. “The speech received at the microphone 134 may be processed by filter 132 and passed to the speech coder 138 for encoding and compression.” *Id.*; *see also id.*, FIG. 5:



Weider “*receive[s] [the] first utterance*” captured through the speech unit. *See, e.g., id.*, [0028] (“For each user utterance, ...the system may perform multiple steps” including “capturing the user’s question or query.”). Weider in combination with Gold therefore renders limitation [1.b(i)] obvious.

[1.b(ii)] “determine, based on processing of the first utterance by a speech recognition engine, one or more words or phrases of the first utterance;”

100. Weider in combination with Gold renders this limitation obvious. Speech input received at Weider’s natural language system, including the “*first utterance*,” is “passed to the speech coder 122 for decoding and decompression.” *Id.*, [0146]. “[T]he output of the speech coder may be fed to the speech recognition engine 120,” which “recognizes words and phrases” and “transcribe[s] utterances to textual messages.” *Id.*, [0187], [0133]; *see also id.*, FIG. 5:



A POSITA would have recognized that Weider's speech recognition engine meets the claimed "*speech recognition engine*" because it is "software or hardware that recognizes the words or phrases in the natural language utterance." EX1021, 2.

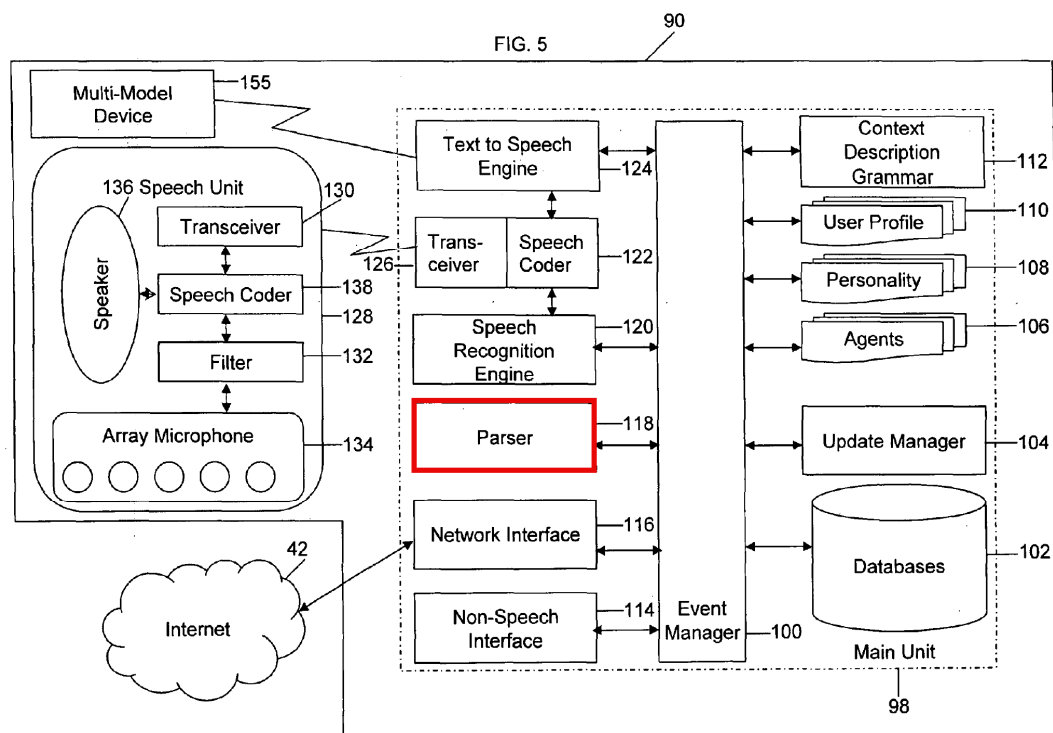
101. Weider discloses that transcription by the speech recognition engine "may be performed using one or more conversational grammars, a context description grammar 112, or other transcription methods." EX1004, [0133]. For example, "[t]he command or request may be compared against a context description grammar to identify a match. Any active grammars in the context description grammar may be scored against the command or request and a best match may be sent to a response generator module." *Id.*, [0020]; *see also id.*, [0117] ("The multi-pass automatic speech recognition module may use a dictation grammar or a large vocabulary grammar... to transcribe the verbal utterance into a text message."). This step is performed for each utterance received at the system. *Id.*, [0028].

102. A POSITA would have understood that, by recognizing words and phrases in the decoded speech and generating a textual transcription of a natural language utterance, Weider's "*speech recognition engine*" performs "*processing of the [] utterance*" to "*determine...one or more words or phrases of the [] utterance.*" *See also id.*, [0146] ("The decoded speech may be processed by the speech recognition engine 120"). Because these steps are performed for each utterance (*id.*, [0028]), the speech recognition engine "*determines one or more words or phrases*

of the first utterance” (the command to play a specific song title). Weider in combination with Gold therefore renders limitation [1.b(ii)] obvious.

[1.b)(iii)] “provide the one or more words or phrases as an input to a conversational language processor;”

103. Weider in combination with Gold renders this limitation obvious. “Once the words and phrases have been recognized by the speech recognition engine 120,” Weider’s speech recognition engine “passes these [words and phrases] to the parser 118 for interpretation.” *Id.*, [0192], [0187]; *see also id.*, FIG. 5:



104. Specifically, Weider discloses that the speech recognition engine passes “tokens” corresponding to the recognized words and phrases to the parser, which “examines the tokens” and determines a context for the utterance “by applying

prior probabilities or fuzzy possibilities to keyword matching, user profile 110, dialog history, and context stack contents.” *Id.*, [0192]. The context “may determine the domain and thereby, the domain agent 156, if any, to be invoked,” and may correspond to the type of command or action to be performed. *Id.* The parser “uses a scoring system to determine the most likely context or domain for a user’s question and/or command,” and “[b]ased on this scoring, the system 90 invokes the correct agent.” *Id.*, [0193].

105. A POSITA would have recognized that Weider’s parser functions as a “*conversational language processor*” by interpreting the recognized words or phrases of the natural language utterance, e.g., by examining the tokens of the recognized words and phrases for context and “applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, dialog history, and context stack contents.” *Id.*, [0192].

106. This is consistent with the disclosure of the ’025 patent, which describes that the conversational language processor “attempt[s] to determine an intent for the multi-modal interaction” by “determin[ing] a most likely context for the interaction from the preliminary interpretation of the utterance.” EX1001, 4:37-42. Moreover, like Weider’s parser, which “invoke[s] the correct agent” for an utterance based on the determined context, the conversational language processor of the ’025 patent “invoke[s] one or more agents 132c having capabilities for

processing requests in a particular domain or application.” EX1004, [0195]; EX1001, 13:9-12. A POSITA would have recognized, therefore, that Weider’s parser meets the claimed “*conversational language processor*.” Additionally, a POSITA would have understood that by “passing” the recognized words and phrases to the parser (conversational language processor) for interpretation, the speech recognition engine “*provide[s] the one or more words or phrases as an input to [the] conversational language processor*.” Weider in combination with Gold therefore renders limitation [1.b(iii)] obvious.

[1.b(iv)] “interpret the first utterance, at the conversational language processor, based on the one or more words or phrases;”

107. Weider in combination with Gold renders this limitation obvious for the reasons I discuss above for limitation [1.b(iii)]. A POSITA would have recognized that Weider’s parser (the “*conversational language processor*”) “*interpret[s] the first utterance...based on the one or more words or phrases*” because, as discussed above, Weider’s speech recognition engine “passes” the recognized words and phrases “to the parser 118 for interpretation.” EX1004, [0187]; *see also id.*, [0028] (“For each user utterance, including a question or query...the system may perform multiple steps possibly including:...parsing and interpreting the question or query.”). Specifically, a POSITA would have understood that the parser uses the “tokens” corresponding to the recognized words and phrases to determine a context for the utterance (e.g., a music context), thereby

“interpreting” the first utterance. See EX1004, [0187] (“The speech recognition engine 120 recognizes words and phrases, using information in the context description grammar 112 and passes these to the parser 118 for interpretation.”). Weider in combination with Gold therefore renders limitation [1.b(iv)] obvious.

[1.b(v)] “identify, based on the interpretation of the first utterance, content to dedicate to a recipient;”

108. Weider in combination with Gold renders this limitation obvious. Weider discloses “*identify[ing], based on the interpretation of the first utterance, content.*” As I discuss for the preamble [1.pre], a user can “command the system to play a specific song title” and the system “provide[s] a song” in response. EX1004, [0139]. A POSITA would have recognized that Weider’s natural language system “*identiffies*” a song (content) in order to provide that song to a user and would have understood that such identifying is “*based on the interpretation of the first utterance*” by Weider’s parser (the “*conversational language processor*”) as I discuss above.

109. To the extent Weider does not explicitly disclose “*identify[ing]... content to dedicate to a recipient,*” the combination with Gold renders this obvious. As I discuss for the preamble [1.pre], Gold’s sender “enters song search criteria 14 such as a song title, a performing artist’s name, or the year a song was first released” and “selects one of the songs from the received list [of search results] to be the dedicated song.” EX1005, [0031]; see also *id.*, [0032], [0036] (“The sender reviews

the song list and selects a song to dedicate to a recipient.”). A POSITA would have understood that Gold’s system “*identif[ies]*” the song (content) when it receives the song title selected by the sender to process the dedication request and initiate the dedication to the recipient.

110. A POSITA would have found it obvious to integrate Gold’s content dedication functionality into Weider’s interactive natural language system. The combined system enables a user to speak into Weider’s natural language speech interface an utterance to select a song title, causing the system to “*identify...content to dedicate to a recipient*” (i.e., the user-specified song). Weider discloses that the user can use the speech interface to select a song title. EX1004, [0139] (“the user may command the system to play a specific song title”). A POSITA would have found it obvious for the user to use the natural language interface to command the system to dedicate the song as I discuss for the preamble [1.pre]. *See id.*, [0004] (“the ability to allow a person to make natural language speech-based requests remains a desirable goal”). A POSITA would have been motivated to make this combination to further supplement the “wide range of domains” with which Weider’s speech interface may be used. EX1004, [0001]. This combination involves application of a known technique (Gold’s content dedication functionality) to improve a similar device (Weider’s interactive natural language system) in the same way (enabling users to dedicate content).

111. Alternatively, as I discuss above for the preamble [1.pre], it would have been obvious to combine Gold and Weider for reasons similar to those the Examiner identified for the Gold-Kennewick combination during prosecution of the '025 patent. EX1003, 187-98 (Final Office Action), 129-44 (Examiner's Answer). Specifically, it would have been obvious to a POSITA to modify Gold by substituting Weider's technique of identifying content from natural language utterances for Gold's UI-based technique for identifying content. A POSITA would have been motivated to make this combination for the reasons I discuss above for the preamble [1.pre].

112. Weider in combination with Gold therefore renders limitation [1.b(v)] obvious.

[1.b(vi)] “initiate, based on the identified content, a dedication to the recipient;”

113. Weider in combination with Gold renders this limitation obvious. Weider discloses “*initiat[ing]*” an action “*based on the identified content.*” For example, as I discuss above for limitation [1.b(v)], Weider's interactive natural language system “provide[s] a song” in response to a user command to “play a specific song title.” EX1004, [0139]. In another example, responding to a user command to “tune in my favorite radio station,” the system can “tune the radio station” or “suggest using an alternative option, such as listening to a CD.” *Id.*, [0027]. A POSITA would have recognized that each of these actions are “*initiate[d]*”

based on the identified content.” For instance, a POSITA would have understood that the song the system “provides” (e.g., plays) is the song (content) that it identifies in limitation [1.b(v)] above.

114. To the extent Weider does not explicitly disclose “*initiat[ing]...a dedication to the recipient*,” the combination with Gold renders this obvious. After the user selects a song title for dedication, Gold’s content dedication system sends the user “instructions for recording the voice message” to be “appended to the dedicated song to create the desired dedication.” EX1005, [0007]. For example, “if the sender’s computer has the appropriate hardware and software, a voice message can be recorded using the sender’s computer.” *Id.*, [0007]; *see also id.*, [0070] (“In one alternative embodiment of the invention, the step of recording a voice message is completed by using a microphone and digitizer sound card connected to the sender host hosting the sender Web client.”). Providing the sender with an opportunity to record a voice message to the recipient through the sender’s computer constitutes “*initiat[ing]...a dedication to the recipient*.” The initiating is “*based on the identified content*” because the dedication is for the “*identified content*.”

115. Given that the combination of Weider and Gold for providing a natural language content dedication service is obvious for the reasons I provide for the preamble [1.pre], a POSITA also would have found it obvious to “*initiate...a dedication to the recipient*” as taught by Gold. For example, a POSITA would have

found it obvious to use Weider's speech interface to provide the user with instructions for recording the dedication voice message. *See, e.g.*, EX1004, [0139] (an agent can prompt a user for information), [0147] (an agent can output speech to a user). A POSITA would have been motivated to combine the references in order to create "a complete natural language verbal...query and response environment." *Id.*, [0005]. This combination involves the application of a known technique (Gold's content dedication functionality) to improve a similar device (Weider's interactive natural language system) in the same way (enabling users to dedicate content). Weider in combination with Gold therefore renders limitation [1.b(vi)] obvious.

[1.b(vii)] "receive a second utterance to be associated with the dedication;"

116. Weider in combination with Gold renders this limitation obvious. Weider discloses that the user may provide "speech annotations" for objects stored by a server. *Id.*, [0120], [0121] ("the annotated objects may be stored on a server side"). For example, the user can provide a speech annotation for a content object that is "stored along with the object[]." *Id.*, [0120]. The object can be "digital media." *Id.*, [0120], claim 23. A POSITA would have recognized a speech annotation as an "***utterance***."

117. Gold, in turn, discloses "***receiving***" an utterance "***to be associated with the dedication***." As I discuss for limitation [1.b(vi)], Gold's song dedication system sends instructions "for recording a voice message to the sender" to allow the sender

to provide a dedication message “to be combined with the dedicated song to create a dedication.” EX1005, [0038], [0032]. The sender may record the voice message “using the sender’s computer.” *Id.*, [0007]; *see also id.*, [0070] (“[T]he step of recording a voice message is completed by using a microphone and digitizer sound card connected to the sender host hosting the sender Web client.”). Gold’s system receives this voice message, which is an “***utterance to be associated with the dedication.***” *Id.*, [0031] (“The voice message when combined with the dedicated song will create a complete dedication.”), [0039], claim 6 (“receiving from the sender a telephone message for use as the voice dedication message”).

118. A POSITA would have found it obvious to use Weider’s natural language speech system to receive the dedication voice message from the user. In the combination, the user uses Weider’s speech interface to provide a “***first utterance***” that the system uses to “***identify...content to dedicate to a recipient,***” as I discuss for limitation [1.b(v)]. A POSITA would have found it obvious for the user to similarly provide, and the system to “***receive a second utterance to be associated with the dedication.***” Weider discloses receiving voice annotations (utterances) for digital media objects. EX1004, [0120], [0121]. Gold discloses receiving a voice message from the dedication sender. EX1005, [0031], claim 6. Weider discloses a system that enables users to annotate a digital media object by voice. EX1004, [0120], [0127], claims 19, 23. Weider does not limit what a user may say in these

speech annotations. EX1004, [0120]. One example of a speech annotation is described in Gold—a dedication voice message. It would have been obvious to a POSITA to try Gold’s dedication message as Weider’s speech annotation because “song dedications can be an effective way to convey and share complex emotions and memories.” EX1005, [0006]. In the combination, the “*second utterance*” is Weider’s speech annotation, and specifically Gold’s dedication voice message. The combinations furthers Weider’s objective to create “a complete natural language verbal...query and response environment.” EX1004, [0005]. Weider in combination with Gold therefore renders limitation [1.b(vii)] obvious.

[1.b(viii)] “determine, based on a processing of the second utterance by the speech recognition engine, one or more words or phrases of the second utterance;”

119. Weider in combination with Gold renders this limitation obvious. As I discuss above for limitation [1.b.(vii)], Weider as modified by Gold renders obvious “*receiv[ing] a second utterance to be associated with the dedication.*” And as I discuss for limitation [1.b.(ii)], the combination renders obvious “*determin[ing], based on processing of the first utterance by a speech recognition engine, one or more words or phrases of the first utterance.*” A POSITA would have found it obvious to perform the same speech recognition processing described above for the second utterance (the speech annotation dedication message) to allow the speech recognition engine to “*determine...one or more words or phrases*” of the dedication message because Weider discloses performing speech recognition processing on

“each user utterance.” EX1004, [0028] (“For each user utterance, ... the system may perform multiple steps possibly including: 1) capturing the user’s question or query through speech recognition operating in a variety of real-world environments.”).

120. Moreover, as I discuss for limitation [1.b(vii)], Weider’s speech annotation of an object is an utterance. *Id.*, [0120] (The system is “configured to allow speech annotations of objects stored thereon.”), [0127] (disclosing “verbal annotations” of objects). To annotate the objects, Weider discloses that the system “transcribe[s] the speech annotations to textual annotations and store[s] the textual annotations with the object.” *Id.*, [0120], [0127], claim 19. A POSITA would have understood that, by transcribing the speech annotation (utterance) to a textual annotation, Weider’s speech recognition engine “*process[es] // the second ...utterance*” to “*determine...one or more words or phrases of the second utterance.*” Weider in combination with Gold therefore renders limitation [1.b(viii)] obvious.

[1.b(ix)] “provide the one or more words or phrases of the second utterance as textual annotations within metadata of the content; and”

121. Weider in combination with Gold renders this limitation obvious. Weider discloses “*provid[ing] the one or more words or phrases of [an] utterance as textual annotations.*” As I discuss for limitation [1.b(vii)], Weider describes that the system “may be configured to allow speech annotations of objects stored thereon,” where the objects include content, such as “digital media.” EX1004,

[0120], [0127]; *see also id.* [0021] (“devices may be configured to allow verbal annotations of objects”), claim 23 (“wherein the object includes digital media”).

122. Moreover, as I discuss for limitation [1.b(viii)], Weider discloses that the system “transcribe[s] the speech annotations to textual annotations and store[s] the textual annotations with the object.” *Id.*, [0120], [0127], claim 19. A POSITA would have understood that Weider’s system transcribes a speech annotation (utterance) by converting “*the one or more words or phrases*” of the speech annotation to a textual representation. *See, e.g.*, EX1004, [0020] (“a multi-pass automatic speech recognition module...transcribes the utterance to a text message”), [0117] (same), [0133] (“the conversational speech analyzer 804 may include one or more speech recognition engines that transcribe utterances to textual messages...data that is entered in a text format may be merged with data that is transcribed to a textual format from the utterance”), claim 1 (“a transcription module that transcribes the speech and the non-speech communications to create a speech-based textual message”).

123. In addition, Weider’s system stores the textual representation of the speech annotation (the transcription of the “*one or more words or phrases*”) as “*textual annotations within metadata of the content.*” *Id.*, [0123] (“the textual annotations and/or textual descriptions that are associated with the objects may be stored as metadata”). The ’025 patent confirms that Weider’s disclosure of storing

annotations as metadata meets “*provid[ing] the one or more words or phrases...as textual annotations within metadata of the content*” as recited by limitation [1.b(ix)]. The ’025 patent specifically relies on Weider’s disclosure for showing how “annotations may be created and associated with the dedicated content,” saying:

In one implementation, verbal annotations, textual annotations, and other types of annotations may be created and associated with the dedicated content using techniques described in co-pending U.S. patent application Ser. No. 11/212,693, entitled “Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions,” filed Aug. 29, 2005, the contents of which are hereby incorporated by reference in their entirety.

EX1001, 29:56-64. Application Ser. No. 11/212,693 is Weider. *See* EX1004, 1 (21).

124. Weider in combination with Gold renders obvious “*receiv[ing] a second utterance to be associated with the dedication*” as I discuss for limitation [1.b(vii)]. Given this combination, a POSITA would have found it obvious to “*provide the one or more words or phrases of the second utterance as textual annotations within metadata of the content*” as recited by limitation [1.b(ix)]. Specifically, a POSITA would have found it obvious to transcribe the second utterance (the speech annotation dedication message) and store the words or phrases of that utterance as textual annotations within metadata of the dedicated content in view of the combined teachings of the references. A POSITA would have been

motivated to make this combination for the reasons I provide above for limitation [1.b(vii)] and because storing the annotation as metadata (as disclosed by Weider) would have provided an efficient way of conveying the dedication text of Gold to the recipient along with the content (i.e., sending the content and the metadata together) in a single message instead of multiple messages that would result in consumption of additional network resources like bandwidth. Weider in combination with Gold therefore renders limitation [1.b(ix)] obvious.

[1.b(x)] “send information to enable the recipient to access the content and the second utterance.”

125. Weider in combination with Gold renders this limitation obvious. In the combined system, the “***content***” is the song identified by the system as I discuss for limitation [1.b(v)]. The “***second utterance***” is the speech annotation dedication message received from the user as I discuss for limitation [1.b(vii)], the words or phrases of which are “***provide[d]...as textual annotations within metadata of the content,***” as I discuss for limitation [1.b(ix)].

126. Gold discloses “***send[ing] information to enable the recipient to access the content and the second utterance***” (i.e., the song and the dedication message). In particular, Gold discloses that, “[a]s one option” for sending the dedication to the recipient, the dedication web server “sends a first email recipient notification 245 to the recipient” to “inform[] the recipient that a dedication has been made by the sender to the recipient.” EX1005, [0008], [0042]. In one embodiment,

the email message contains a URL the recipient can invoke to access the content and the dedication voice message. *Id.*, [0043] (“The recipient...email notification[] contain[s] the URL of a play dedication server script located on the dedication Web server....The play dedication server script can be invoked by selecting its URL from the body of each email notification...The play dedication server script executes and returns a Web page and a dedication client script...The dedication client script contains the URLs of the voice message and dedicated song.”).

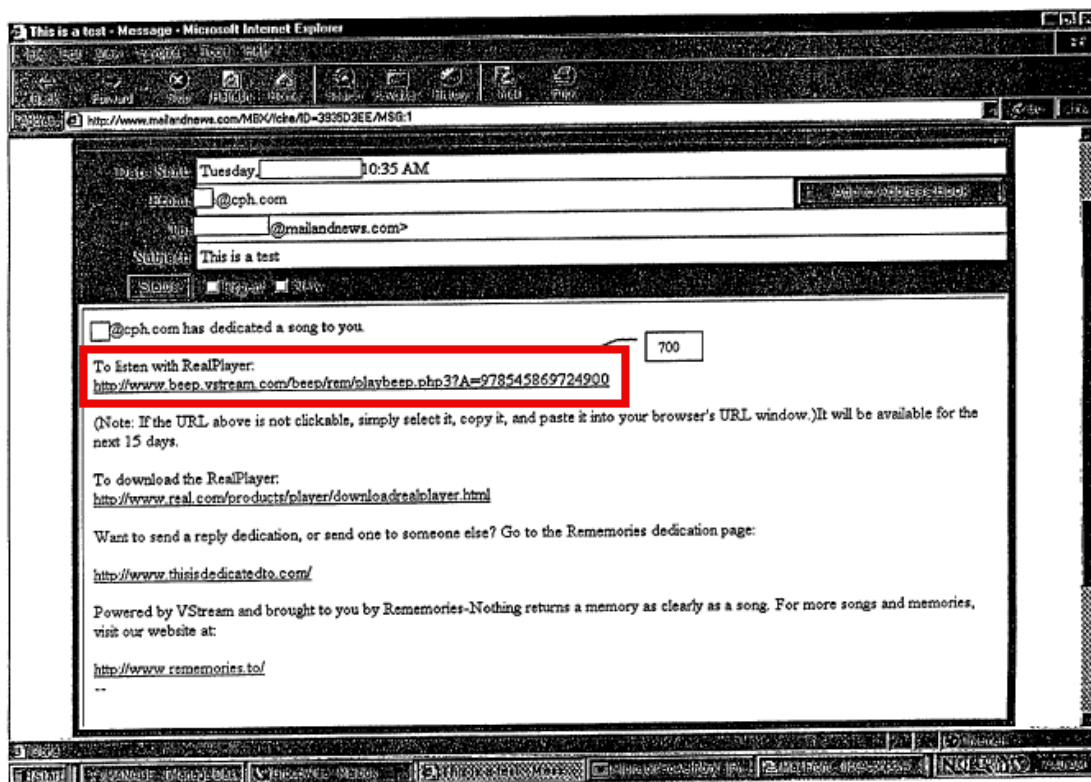


FIG. 11

Id., FIG. 11, [0057] (“Selection of the URL hypertext link 700 invokes the recipient’s Web client and invokes the previously described play dedication server script.”). In another embodiment, the email includes a digital audio file containing

the content and the second utterance as an attachment. *Id.*, [0070] (“the voice dedication and song are combined into a single audio file, preferably in digital format, that is sent in its entirety as an attachment to the email dedication notification sent to the recipient”). In still another implementation, Gold discloses that “the computer telephony system is used to place a telephone call to the recipient to play the dedication.” *Id.*, [0008]. Alternatively, “the dedication and song are made available as a voicemail message retrievable using a telephone.” *Id.*, [0070]. Each of the email, phone call, and voicemail message contains “**information**” that Gold sends to the recipient.

127. In the combination, the one or more words or phrases of the second utterance are “***provid[ed]...as textual annotations within metadata of the content.***” A POSITA would have found it obvious to include the metadata of the content with the content, so that the recipient accessing the content also accesses the metadata. As I discuss for limitation [1.b(ix)], Weider discloses storing textual annotations within metadata of content. EX1004, [0123] (“the textual annotations and/or textual descriptions that are associated with the objects may be stored as metadata”). Weider also discloses that metadata may be included within the content. *Id.*, claim 28 (“wherein the object includes metadata”). Therefore, a POSITA would have found it obvious to provide the metadata of the content to the recipient along with the content by including the metadata within the content or by otherwise sending it

at the same time, thereby consuming less network resources than sending multiple messages. A POSITA also would have recognized that bundling metadata and content was a well-known practice in digital media systems, improving efficiency by reducing the need for separate storage, synchronization, or retrieval operations.

128. A POSITA would have found it obvious for the combined system of Weider and Gold to inform the recipient of the dedication and “*enable the recipient to access the content and the second utterance*” as taught by Gold. The dedication serves as a notification to the recipient. EX1005, [0006] (“[S]ong dedications can be an effective way to convey and share complex emotions and memories when it might not be possible, or desired, for the sender and recipient to communicate directly.”). In the combined system, the user can provide Weider’s interactive natural language system with contact information for the recipient, e.g., by voice, and the combined system sends a dedication to the recipient using the contact information, e.g., by sending an email or phone call or otherwise sending a message using the appropriate contact information. This combination is a combination of prior art elements according to known methods (Weider collecting user input via the speech interface and Gold prompting a user to provide recipient contact information) to yield predictable results (using voice input to collect recipient contact information). A POSITA would have had a reasonable expectation of success because the technologies involved (e.g. sending an email with a URL or attachment,

placing a phone call, or otherwise sending a message to a recipient) were well understood. Weider in combination with Gold therefore renders limitation [1.b(x)] obvious.

2. Claim 3: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to annotate the content with the second utterance.”

129. Claim 3 is substantively identical to limitation [1.b(ix)] and is rendered obvious for the same reasons.

3. Claim 6

[6.a] “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: send, to the recipient, an electronic link to access the content as the information to enable the recipient to access the content; and”

130. Weider in combination with Gold renders this limitation obvious. As I discuss for limitation [1.b(x)], Gold’s dedication Web server “informs the recipient that a dedication has been made by the sender to the recipient.” EX1005, [0042]. Gold further discloses “*send[ing], to the recipient, an electronic link to access the content as the information to enable the recipient to access the content.*” For example, Gold discloses that the information can take the form of a “first email recipient notification,” which “contain[s] the URL of a play dedication server script located on the dedication Web server” (*id.*, [0042], [0043]), as illustrated below in FIG. 11:

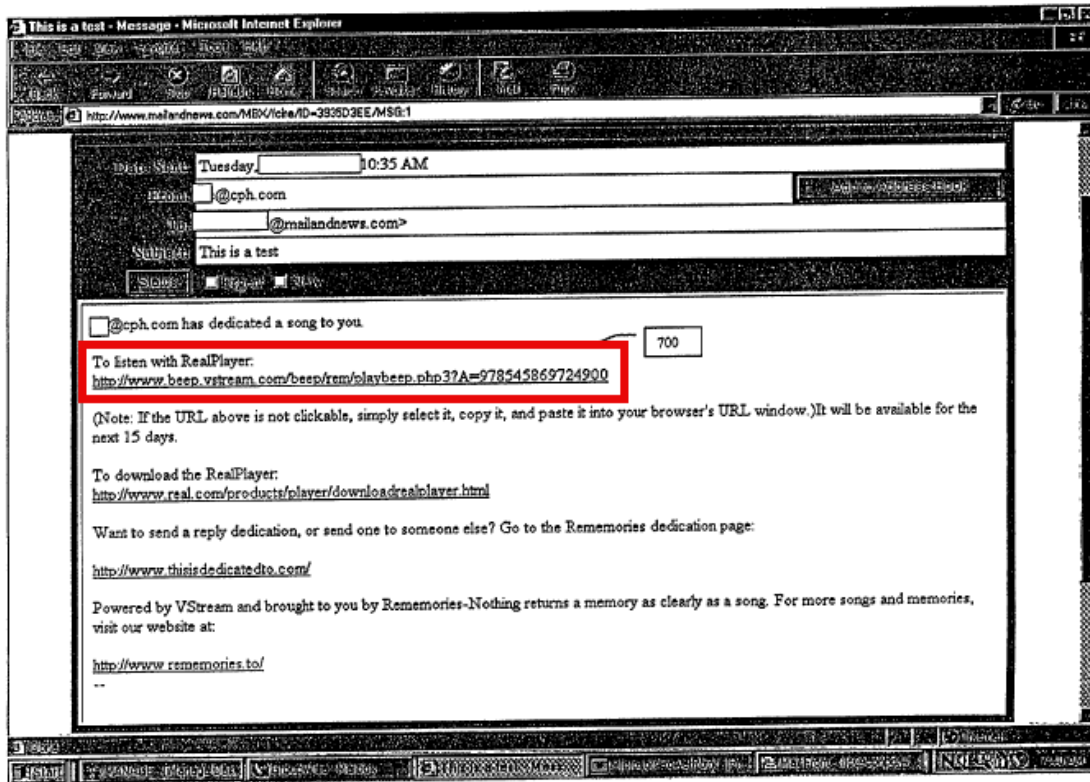


FIG. 11

see also id., [0057] (“Selection of the URL hypertext link 700 invokes the recipient’s Web client and invokes the previously described play dedication server script.”). A URL is an electronic link, and Gold’s dedication Web server “*send[s], to the recipient, [the] electronic link*” (e.g., in the recipient notification email).

131. A POSITA would have understood that the link (e.g., the URL of the play dedication server script) “*enable[s] the recipient to access the content.*” When executed (e.g., when the recipient clicks on the link), Gold’s play dedication server script returns a “dedication client script” that “includes the URLs for the voice message and the dedicated song” and “constitutes the software embodiment of the complete dedication.” *Id.*, [0043]; *see also id.*, [0058]-[0061] (reciting an example

dedication client script including a “URL of the song to be played after the voice dedication”). The recipient Web client executes the dedication client script to “play the voice message and the dedicated song as a single dedication,” thereby enabling the recipient to “access” the song (content). *Id.*, [0043]; *see also id.*, [0056] (“The recipient Web client plays 288 the song for the recipient.”). Weider in combination with Gold therefore renders this limitation obvious.

[6.b] “send, to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance.”

132. Weider in combination with Gold renders this limitation obvious. Weider in combination with Gold renders obvious “*receiv[ing] a second utterance to be associated with the dedication*,” as I discuss for limitation [1.b(vii)]. The combined Weider-Gold system receives the utterance as audio. *See, e.g.*, EX1004, [0120] (receiving “speech annotations” that are transcribed to text); EX1005, [0038] (the sender “create[s] a recorded voice message 240 in much the same way a conventional telephone answering machine is used to leave a voice message”). Weider discloses that the transcription of the speech annotation is “stored as metadata” of the object. EX1004, [0123]. Weider in combination with Gold additionally renders obvious “*send[ing] information to enable the recipient to access the content and the second utterance*,” as I discuss for limitation [1.b(x)].

133. A POSITA would have found it obvious to “*send, to the recipient, the second utterance is an...audio format*.” *See, e.g.*, EX1005, [0008] (“As another

option, the computer telephony system is used to place a telephone call to the recipient to play the dedication.”); [0050] (“[A] phone number 511 of the recipient is requested for sending a song dedication to the recipient using a telephone.”); [0063]-[0064]. For example, it would have been obvious to send an audio dedication along with audio-based content because the user is consuming the content via audio and it would have been logical to make the dedication (second utterance) also available by audio.

134. Additionally, a POSITA would have found it obvious for the combined system of Weider and Gold to “*send, to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance.*” As I discuss for limitation [1.b(x)], “as one option” for sending the dedication to the recipient, Gold’s dedication web server “sends a first email recipient notification 245 to the recipient” to “inform[] the recipient that a dedication has been made by the sender to the recipient.” EX1005, [0042]. The email includes a URL the user can invoke to access the content and the dedication voice message (*id.*, [0043]) or includes a digital audio file containing the content and the voice message as an attachment (*id.*, [0070]).

135. When the email includes a URL, the recipient can access the URL to ultimately “play the voice message and the dedicated song as a single dedication.” *Id.*, [0043]; *see also id.*, [0056] (“The recipient Web client retrieves the voice

message 285 from the dedication Web server using the voice message URL” and “plays the voice message 284 for the recipient.”), [0058]-[0061] (reciting an example dedication client script including a “URL of the voice dedication”). By following the URL in the email, the recipient “access[es] the second utterance” because execution of the dedication client script causes the recipient Web client to “play the voice message.” *Id.*, [0043]. When the email includes the digital audio file as an attachment, a POSITA would have recognized that the recipient accesses the second utterance (dedication message) by opening the attachment. *Id.*, [0070].

136. In both cases, Gold discloses “*send[ing], to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance.*” A POSITA would have understood that computerized systems commonly exchange audio information in “*encoded audio format[s].*” For example, Gold discloses that the file containing the combined song and voice dedication can be in the MP3 format, which a POSITA would have recognized as an encoded audio format. *Id.*, [0070] (“[t]he attachment format can be of any suitable format such as MP3”); *see also, id.*, [0039] (“In the preferred embodiment, the digitized voice message is in the same format as the song so that the song and the digitized voice dedication can be played by the same software application and hardware.”); EX1024, 349 (defining MP3 as “a digital audio coding scheme used in distributing recorded music over the Internet.”). In another example,

where “the computer telephony system is used to place a telephone call to the recipient to play the dedication,” (EX1005, [0008]), a POSITA would have recognized that the second utterance (the speech annotation dedication message) is likewise provided in an encoded audio format. For example, a POSITA would have understood that audio is encoded for transmission over telecommunications networks to reduce bandwidth usage and enable compatibility across devices and networks. Weider in combination with Gold therefore renders this limitation obvious.

4. Claim 7: “The system of claim 1, wherein the content includes one or more of a song or a video.”

137. Weider in combination with Gold renders claim 7 obvious. Gold’s system is a “song dedication system” by which a user can “send dedications of popular songs through a communications medium such as the Internet.” EX1005, [0010], Abstract. Alternatively, “a music video may be dedicated by the sender instead of a song.” *Id.*, [0070]. A POSITA would have understood and found obvious that the combined system of Weider and Gold provides “*one or more of a song or a video*” as the dedicated content for the reasons I provide for claim 1. Weider in combination with Gold therefore renders claim 7 obvious.

5. Claim 8

138. Weider in combination with Gold renders claim 8 obvious for similar reasons as claim 1. As I discuss above for the preamble [1.pre] and limitations [1.a]-

[1.b], Weider in combination with Gold renders obvious “[a] *system for providing a natural language content dedication service*,” where the system comprises “*one or more processors*” and “*one or more non-transitory computer readable mediums storing executable instructions that when executed by the one or more processors cause the one or more processors to*” perform programmed operations.

139. A POSITA would have recognized that the combined Weider-Gold system performs “*a method for providing a natural language content dedication service*” for the same reasons. For example, Weider is titled “Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions” and discloses that its interactive natural language system performs a “method of processing speech and non-speech communications.” EX1004, claim 38. Similarly, Gold is directed to a “system and method for using the Internet or electronic media employing computerized database approaches for tying together songs and personal voice messages to send song dedications.” EX1005, [0007]; *see also id.*, claim 1 (“A method of sending a song dedication from a sender to a recipient.”). A POSITA would have understood that the systems of Weider and Gold, which I discuss above for the preamble [1.pre] and limitations [1.a]-[1.b(x)], perform the described methods, including “*a method for providing a natural language content dedication service*.” The remaining limitations of claim 8 are substantively identical to those

of claim 1 and are obvious for the reasons I provide above for claim 1. Weider in combination with Gold therefore renders claim 8 obvious.

6. Claim 10: “The method of claim 8, further comprising annotating the content with the second utterance.”

140. Claim 10 is substantively identical to claim 3 and is rendered obvious for the same reasons.

7. Claim 13: “The method of claim 8, further comprising: sending, to the recipient, an electronic link to access the content as the information to enable the recipient to access the content; and sending, to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance.”

141. Claim 13 is substantively identical to claim 6 and is rendered obvious for the same reasons.

8. Claim 14: “The method of claim 8, wherein the content includes one or more of a song or a video.”

142. Claim 14 is substantively identical to claim 7 and is rendered obvious for the same reasons.

9. Claim 19: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide the second utterance within the content.”

143. Weider in combination with Gold renders claim 19 obvious. Specifically, Gold discloses that, “in an alternative embodiment, the voice message and the song are combined to create a single electronic data file suitable for replay by a digital audio system.” EX1005, [0040]. “This single electronic data file is sent to a recipient and the single electronic data file is retained by the recipient for

replay.” *Id.*; see also EX1005., [0070] (“In still another embodiment, the voice dedication and song are combined into a single audio file, preferably in digital format, that is sent in its entirety...to the recipient.”). A POSITA would have recognized that, where the dedication (second utterance) is provided within the same file as the song (content), the dedication is provided “*within the content.*” Weider in combination with Gold therefore renders claim 19 obvious.

B. Ground 2: Weider in view of Gold and Martinez renders obvious claims 2, 9, 17, and 18

1. Claim 2: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide an image as one or more tags within metadata of the content.”

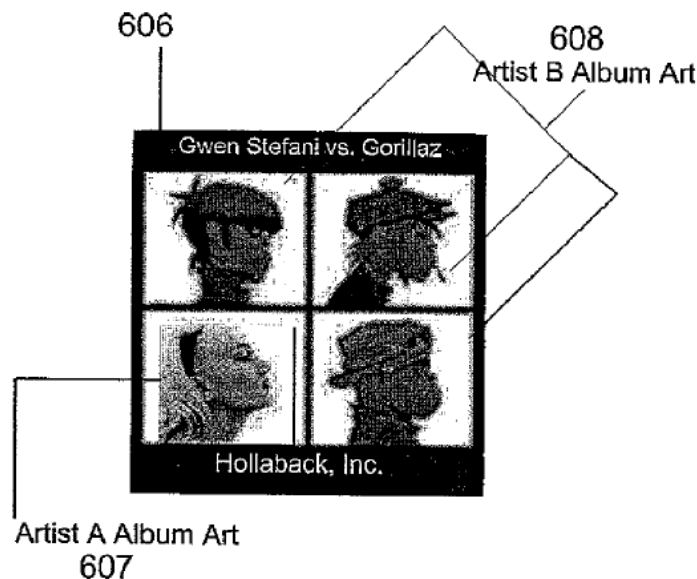
144. Weider in combination with Gold and Martinez renders claim 2 obvious. As I discuss above for limitation [1.b(ix)], Weider in combination with Gold renders obvious “*provid[ing] the one or more words or phrases of the second utterance as textual annotations within metadata of the content.*” Martinez further discloses “*provid[ing] an image as one or more tags within [the] metadata of the content.*”

145. Martinez describes a container-based approach for storing content metadata. For example, Martinez discloses a “set of metadata for association with a composite media item,” where the metadata is “inserted...into at least one data container” (EX1006, [0032], [0020]) comprising “any number and type of set(s) of

metadata associated with a media item portion, any number and type of set(s) of metadata associated with the container itself and any number and type of media item identifier(s).” *Id.*, [0035]. A POSITA would have recognized that data containers were a well-known file format for encapsulating both media data (e.g., audio data in an MPEG file) and associated metadata, such as text-based tags that describe the content (e.g., song title, artist, URL, or links to other related information or data). The tags may include, for example, license information, pointers to file locations, content identifying information, etc. *See id.*, [0050].

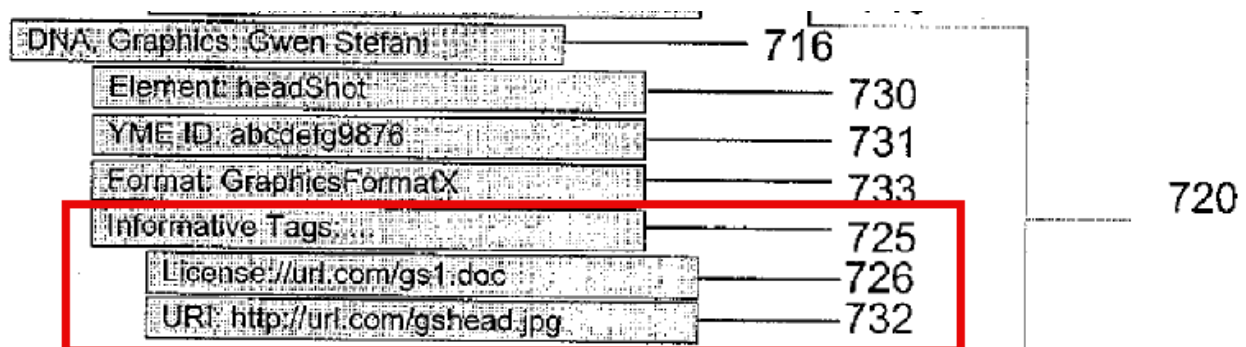
146. Martinez explicitly discloses using such tags within data containers to store metadata, reciting that the data container “comprises at least one set of informative tags associated with the composite media item.” *Id.*, [0016]; *see also id.*, [0050] (“[T]he informative tags...are associated with the metadata subsets.”). A container “may be created using a variety of methods such as XML, using linear strings of data with internal header information, or software objects such as those utilized in the Java or C++ programming language.” *Id.*, [0044]; *see also id.*, [0004] (“[M]etadata is typically entered...into whatever tags are available, such as, for example, ID3 tags in MP3 files.”). A POSITA would have recognized that these methods, including XML and linear strings of data (e.g., ID3 tags), utilize tags to structure information.

147. Martinez further discloses “*provid[ing] an image as one or more tags within metadata of the content.*” For example, Martinez acknowledges that metadata associated with a media item may include an image, e.g., a “graphical image such as a jpeg.” EX1006, [0045]. A POSITA would have understood that media content (e.g., songs) typically have associated images, such as album cover art. For example, in the case of a composite media item “Hollaback Inc.,” Martinez explains that the associated metadata includes “a graphic image 606” that is “created from the combination of two other metadata subsets; a graphic image from ‘Hollaback’ 607 and a graphic image from ‘Feel Good Inc’ 608.” *Id.*, [0058]; *see also id.*, FIG. 6 (excerpted):



see also id., FIG. 7. As illustrated in FIG. 6, the graphic images 607 and 608 are “album art.”

148. A POSITA would have found it obvious to provide an image (e.g., album cover art) as a “*tag[] within metadata of the content.*” See, e.g., *id.*, [0049]-[0050] (“[T]he informative tags...are associated with the metadata subsets” including a metadata subset comprising “static media such as a JPEG still image”). Martinez describes that a metadata tag in a data container could be used to point to a location of an image or contain the image data itself. *Id.*, [0052], [0058]. For example, Martinez illustrates an image metadata tag in the following excerpt of FIG. 7, in which metadata subset 720 is “a graphical image such as a jpeg.” *Id.*, [0045]. The metadata subset 720 includes an informative tag 732, which “comprises a pointer to a physical file location where a media item portion or media item may be found[:.]”



Id., [0052], FIG 7. Alternatively, “the media item portion” “might be directly embedded within the metadata subset and/or data container.” *Id.*, [0052]. Because the image data (media item portion) or a reference to an image can be stored as the value of a tag, the image is “*provide[d] as one or more tags within metadata of the content.*”

149. It would have been obvious to a POSITA to modify the combined Weider-Gold system to include the image metadata of Martinez. For example, a POSITA would have found it obvious for the content dedication service to provide an image (e.g., album art) with a dedicated song to allow the recipient to more easily identify the song, the artist, and/or the album on which the song is included. *See, e.g., id.*, [0038] (a media item may include “a jpeg image media item portion”); *id.*, [0041] (“[T]he media item portion may be any type of identifiable portion or component of media item,” such as “a graphical image”). Providing the image as a ***“tag[] within metadata of the content”*** would have provided an efficient way of conveying the image to the recipient along with the content (i.e., sending the content and the image metadata together), e.g., by keeping the image and content packaged together, improving playback and display consistency, and simplifying the system architecture.

150. This is a combination of prior art elements according to known methods to yield predictable results. Both Weider and Martinez acknowledge that metadata may include many different types of data. *See, e.g.,* EX1004, [0123] (“One of ordinary skill in the art will readily appreciate that metadata may include various types of information.”); EX1006, [0047] (“[A] metadata subset may comprise any number and combination of information or data about a media item portion”). One such data type disclosed in Martinez includes image tags. EX1006, [0045] (metadata

may include a “graphical image such as a jpeg”). Combining Martinez’s image metadata with the Weider-Gold system that permits different metadata types yields the predictable result of the metadata including the image tags of Martinez. Weider in combination with Gold and Martinez therefore renders claim 2 obvious.

2. Claim 9: “The method of claim 8, further comprising providing an image as one or more tags within metadata of the content.”

151. Claim 9 is substantively identical to claim 2 and is rendered obvious for the same reasons.

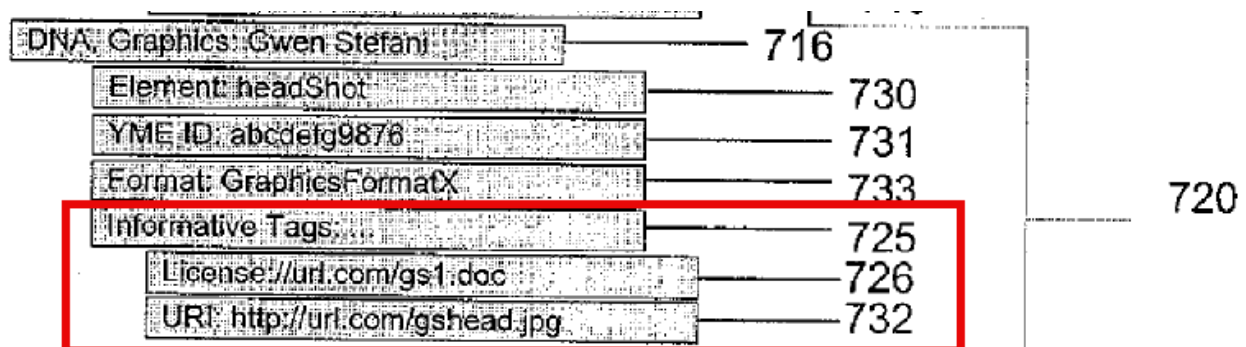
3. Claim 17: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide an electronic link associated with the content as one or more tags within metadata of the content.”

152. Weider in combination with Gold and Martinez renders claim 17 obvious. As I discuss above for claim 2, Weider in combination with Gold and Martinez renders obvious “*providing [data] as one or more tags within metadata of the content.*”

153. Martinez further discloses that the data includes “*an electronic link associated with the content.*” For example, Martinez broadly defines the metadata subset as “compris[ing] information or data about a media item portion,” such as “a URL or link to the official webpage of the artist, album, rights holder, or producer associated with the media item.” EX1006, [0037] [0047]; *see also id.*, [0040]. A

URL is an “*electronic link*,” and each of the listed webpages is “*associated with the content*.”

154. Additionally, Martinez discloses “*provid[ing]*” the link “*as one or more tags within metadata of the content*.” For example, as illustrated in the following excerpt of FIG. 7, the metadata subset 720 of FIG. 7 includes informative tags 725:



Martinez discloses that informative tag 726 comprises “a pointer to licensing information” associated with the media item while informative tag 732 “comprises a pointer to a physical file location where a media item portion or media item may be found.” *Id.*, [0051], [0052]. Each of these pointers is a tag within the metadata (metadata subset 720) of the content (the media item) and comprises “*an electronic link associated with the content*.”

155. It would have been obvious to a POSITA to modify the combined Weider-Gold system to “*provide an electronic link associated with the content as one or more tags within metadata of the content*,” as disclosed by Martinez. For example, where the metadata includes an image (e.g., album art), a POSITA would

have recognized that providing a pointer or an electronic link to the image within the metadata would have reduced the metadata file size and enabled faster load and transfer times than if the metadata contained the image data itself. A POSITA would have been motivated to make this combination for the reasons I provide above for claim 2. Weider in combination with Gold and Martinez therefore renders claim 17 obvious.

4. Claim 18: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide a voice tag as one or more tags within metadata of the content.”

156. The '025 patent describes a “*voice tag*” as a natural language utterance. EX1001, 29:22-29 (“a natural language utterance that includes information to be inserted as voice-tags for the dedicated content”), 29:52-56 (“any utterances to be inserted as voice-tags for the dedicated content may provide further verbal annotations for the dedicated content”).

157. Weider in combination with Gold and Martinez renders claim 18 obvious. As I discuss above for claim 2, Weider in combination with Gold and Martinez renders obvious “*providing [data] as one or more tags within metadata of the content.*”

158. Martinez discloses that the metadata tag may be in audio format. Martinez explains that metadata “may comprise any number and combination of information or data about a media item portion,” including “audio.” EX1006,

[0047]; [0049]; *see also id.*, [0045] (“[T]he second metadata subset 719 is the A cappella vocals [of the media item], that is, the same vocals as included in the complete media item, but without any instrumentals.”).

159. It would have been obvious that the audio metadata tag could take the form of a natural language utterance (voice tag) e.g., to allow the sender of the dedication to provide any additional natural language utterances to be associated with the dedicated content. It would have further been obvious to “*provide [the] voice tag[s] as one or more tags within metadata of the content*” in view of Martinez’s disclosure of tag-based methods for creating data containers. *Id.*, [0044] (“[S]uch containers may be created using a variety of methods such as XML, using linear strings of data with internal header information”). A POSITA would have been motivated to make this combination for the reasons I provide above for claim 2. Weider in combination with Gold and Martinez therefore renders claim 18 obvious.

C. Ground 3: Weider in view of Gold and Ginter renders obvious claims 4 and 11

1. Claim 4: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to purchase the content from a content provider to enable the recipient to access the content.”

160. Weider in combination with Gold and Ginter renders claim 4 obvious. As I discuss for the preamble [1.pre], Gold’s content dedication system includes a

“database of songs for use in the dedication.” EX1005, [0007]. The database “is preferably searchable by criteria such as title, artist, and year the song was released.”

Id. In response to receiving a user’s song list request, Gold’s jukebox Web server “searches the...available database of songs for songs matching the song list form data” and “compiles a song list” that it returns to the user Web client. *Id.*, [0036].

161. To the extent Gold does not explicitly disclose that the content dedication service “***purchase[s] the content***” for the song database “***from a content provider,***” the combination with Ginter renders this obvious. Ginter discloses a “‘virtual distribution environment’ (called VDE...) that secures, administers, and audits electronic information use” by “managing content that travels ‘across’ the ‘information highway.’” EX1009, 2:19-24. The VDE can “protect the rights of parties who have...commercial interests in electronically distributed information,” such as content creators and distributors, by “ensur[ing], for example, that parties, will be paid for use of distributed information in a manner consistent with their agreement.” *Id.*, 4:28-34. “For example, an electronic agreement between a content creator and a distributor may establish...the price to the distributor for a creator’s content...that this distributor may distribute to end users.” *Id.*, 47:24-30. A POSITA would have understood that Ginter’s distributor purchases the content from the content creator, e.g., according to the terms specified in the electronic agreement, because such is the purpose of the electronic agreement (e.g., the purchase enables

the distributor to distribute legally the content to end users). *See also id.*, 7:58-63 (“Electronic rights protection provided by the present invention will also provide an important foundation for...electronic purchasing”).

162. It would have been obvious to modify the combined Weider-Gold system to incorporate Ginter’s disclosed method for purchasing content items from content creators. Ginter discloses that the purchased content includes music. *See, e.g., id.*, 1:55-61 (“The owners and marketers of this content include...recording companies”), 56:48-51 (“[A] musical performing group may act as both content creator 102 and distributor 106 by creating and distributing its own musical recordings.”), 59:28-32 (“Information content 304 could be...sound such as a musical performance”), 141:8-14 (“[C]ontent may include:...music”). Gold discloses that the content dedication service includes a song database. EX1005, [0007]. Likewise, Ginter discloses that the purchased content “may be stored in content storage” of a repository for distribution to end users. EX1009, 314:1-2, FIG. 78. Ginter’s repository “includes a content catalog...[of] content available through the repository,” including “identifying information” and “the charge for certain content usage activities.” *Id.*, 314:24-52.

163. In the combined system, Gold’s service functions as a distributor, and the content creator of Ginter is the content provider. *See, e.g.,* EX1009, 3:48-59 (“[T]he present invention...can help ensure that content providers will be paid for

use of distributed information.”), 2:57-59 (“Commercial content providers are concerned with ensuring proper compensation for the use of their electronic information.”); *see also id.*, FIG. 80.

164. It would have been obvious for Gold to populate the song database by **“purchas[ing] the content from a content provider”** (as disclosed by Ginter) because this is how music is conventionally acquired, i.e., users and services buy the right to use songs from the rights holders. *See, e.g.*, EX1016, [0010] (“[T]o obtain endorsement [by the music industry], companies must charge users for access to songs.”), [0051] (A user “engage[s] in an on-line commerce transaction in which the user pays for access rights to [a] particular media item.”); EX1009, 3:19-21 (“VDE allows the owners and distributors of electronic digital information to reliably bill for, and securely control, audit, and budget the use of, electronic information.”). A POSITA would have been motivated to make this combination because, for example, Ginter discloses storing purchased content in a content repository for distribution to users. EX1009, 309:25-38, 311:44-51, 314:1-2, 315:59-316:15, FIG. 78.

165. In the combination, the Weider-Gold-Ginter content dedication system purchases content from content providers (e.g., artists or record companies) and functions as a distributor by sending (distributing) information **“to enable the recipient to access the content,”** as I discuss above for limitation [1.b(x)]. The use

of Ginter’s disclosed method of purchasing content to build the song database of Gold is a combination of prior art elements according to known methods – Gold’s distribution of database content to users (dedication recipients) and Ginter’s method of acquiring content via purchase from content creators – to yield predictable results – a song dedication system populated with content purchased from content creators. The result of the combination is predictable because conventional business relationships include sales transactions. A POSITA would have reasonably expected the combination to be successful for the same reasons. Weider in combination with Gold and Ginter therefore renders claim 4 obvious.

2. Claim 11: “The method of claim 8, further comprising purchasing the content from a content provider to enable the recipient to access the content.”

166. Claim 11 is substantively identical to claim 4 and is rendered obvious for the same reasons.

D. Ground 4: Weider in view of Gold, Ginter, and Wang renders obvious claims 5 and 12

1. Claim 5

[5.a] “The system of claim 4, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: identify, for the dedication, a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option; and”

167. I understand the plain-and-ordinary meanings of the “*buy-to-own*” and “*pay-to-play*” purchase options in this claim limitation to refer to purchase of the

content associated with the dedication. This is consistent with the disclosure of the '025 patent, which recites that “*buy-to-own*” includes “purchasing full rights to the requested content” and that “*pay-to-play*” includes “purchasing rights from the content provider that permits the purchased content to be played a predetermined number of times.” EX1001, 35:41-61. I also understand that the plain-and-ordinary meaning of the “*paid subscription purchase option*” refers to a subscription for use of the content dedication system, which is consistent with the '025 patent disclosure that “the user may pay a periodic service charge to the content dedication system that permits the user to ... make content dedications in accordance with terms defined in a subscription.” *Id.*, 35:62-36:7.

168. Weider in combination with Gold, Ginter, and Wang renders this limitation obvious. Gold discloses that the dedication service may allow senders to “use the invention to purchase and send a dedication for a fee.” EX1005, [0009]. However, to the extent Gold does not explicitly disclose that the fee corresponds to “*a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option,*” the combination with Wang renders this obvious.

169. Wang discloses that a user can purchase an identified song. As its title specifies, Wang is directed to a “Method and System for Purchasing Pre-Recorded Music.” Wang describes a “signal identification service” that enables a user to

identify and purchase a song through an “interactive voice response unit (‘IVR’).” EX1007, [0012], [0070]. The service may provide the user with multiple “purchase options” for the identified music. *Id.*, [0026], [0084]. For example, the user can download a digital copy of a purchased song. *Id.*, [0025] (“[D]igital delivery and fulfillment techniques may also be used, including digital downloads to users’ PCs, wireless internet devices, hand held personal computers palm pilots, or mobile telephones.”). A POSITA would have found it obvious that a purchase including a digital download could be a “*buy-to-own purchase option*” because the user’s purchase entitles it to own the digital file it downloaded. Wang also contemplates subscription-based options for use of the signal identification service, disclosing that “subscription arrangements with users may be established and revenue received.” *Id.*, [0062]; *see also id.*, [0038] (“Service-specific information may also be furnished...to the service subscribers on a given day.”). A POSITA would have recognized that a “subscription arrangement[]” that generates revenue for use of the signal identification service is a “*paid subscription purchase option*.” Wang therefore discloses “*one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option*.”

170. Wang also discloses “*identify[ing],...a purchase option*.” When the system identifies a song, “the user is then provided the opportunity to immediately place an order for a recording of the identified music,” via the disclosed purchase

options, such as by purchasing “an album on CD or tape that contains the selected track,” or “digital download[.]” *Id.*, [0024]-[0025]. According to Wang, “fulfillment of the desired transaction to the user can be accomplished by the service directly.” *Id.*, [0049]. Specifically, Wang discloses that “billing data from user billing block 125...may be used to generate a bill to the user commensurate with the extent of the usage of the signal identification service.” *Id.*, [0070]. Because Wang provides multiple purchase options for the identified song and bills the user “commensurate with the extent of the usage” of the service, a POSITA would have understood that the signal identification system could provide the different purchase options to the user and “*identify...a purchase option*,” in response to the user selection, e.g., to determine an amount to bill the user.

171. A POSITA would have found it obvious to modify the combined Weider-Gold-Ginter system to incorporate the disclosed purchase options taught by Wang. First, it would have been obvious to implement Wang’s “buy-to-own” purchase option for the dedication within the Weider-Gold-Ginter system to allow a recipient of a dedication to download the dedicated song and replay the song on command. This would have been an obvious way to implement the embodiment of Gold in which the system provides the dedication (including the dedicated song) in a “physical computer file” (e.g., as an attachment to the email notification). EX1005, [0008]. Second, Gold discloses that “senders may use the invention to purchase and

send a dedication for a fee.” EX1005, [0009]. It would have been obvious for the fee to take the form of a paid subscription for use of Gold’s content dedication service, as taught by Wang or a “buy-to-own” fee, e.g., for each dedication the user sends because both “*buy-to-own*” and “*paid subscription*” purchase models for content and use of services were well known. *Id.*, [009]. Moreover, like Gold, Wang contemplates that the user can share content with others, such as by “instruct[ing] the service to send a recommendation for the identified song to a friend via email or SMS or a clip of the song to the friend’s mobile telephone.” EX1007, [0043]; *see also id.*, [0042]. The result of the combination is predictable for the same reasons, and A POSITA would have reasonably expected the combination to be successful.

172. Additionally, as I discuss above for claim 4, Weider in combination with Gold and Ginter renders obvious “*purchas[ing] the content from a content provider to enable the recipient to access the content.*” Where the identified purchase option is a “*buy-to-own purchase option,*” it would have been obvious to a POSITA that a portion of the user payment would be passed to the content provider as part of the payment that the distributor (Gold’s service) makes to the content provider (Ginter’s content creator) to purchase the content. This is consistent with the disclosure of the ’025 patent, which recites that the amount charged to the user “encompasses the cost for purchasing the content from the content provider 680 plus a service charge for dedicating the content.” EX1001, 30:48-58. Alternatively,

where the identified purchase option is a “*paid subscription purchase option*,” it would have been obvious for the combined system to use a subscription service to recuperate the cost of building the song database, as I discuss above for claim 4.

173. A POSITA would have recognized that the various disclosed purchase options would have enabled the operator of the combined system to generate profit for providing a service and would have found it obvious to incorporate at least one purchase option into the combined Weider-Gold-Ginter system because each is a well-known billing model. Faced with a need to charge users (EX1005, [0009]), a POSITA would have looked to known billing strategies for content delivery, as disclosed by Wang. *See, e.g.*, EX1016, [0010] (“[T]o obtain endorsement [by the music industry], companies must charge users for access to the songs.”), [0051] (A user “engage[s] in an on-line commerce transaction in which the user pays for access rights to [a] particular media item.”). This combination includes use of a known technique (Weider’s digital download or subscription-based monetization model) to improve a similar product (a digital content system) in the same way (enabling revenue generation). Weider in combination with Gold, Ginter, and Wang therefore renders limitation [5.a] obvious.

[5.b] “bill a user associated with the first utterance based on the purchase option.”

174. Weider in combination with Gold, Ginter, and Wang renders this limitation obvious. First, Wang discloses “*bill[ing] a user*.” A user may “instruct

the service to send a recommendation for the identified song to a friend via email or SMS,” and the system “generate[s] a bill to the user commensurate with the extent of the usage of the signal identification service.” EX1007, [0043], [0070] *see also id.*, [0027] (“[A]fter [a] song selection is identified, a sales order is automatically placed. The user’s account is then debited, and the selection is forwarded to the user.”).

175. In the combined system, the user who makes the content dedication request is billed for the dedication as taught by Gold, and the fee this user is billed is based on the identified purchase option as taught by Wang. Because the user selects a song title by speaking the first utterance, the user who makes the request is ***“associated with the first utterance.”*** It would have been obvious that the user who is billed for the content dedication (according to the identified purchase option) is the user who is ***“associated with the first utterance.”*** For example, a POSITA would have recognized that the user ***“associated with the first utterance”*** is the user initiating the transaction and sending the dedication, and it would have been obvious to bill the sender to avoid the recipient incurring an unexpected charge.

176. It would have been obvious to a POSITA to modify the Weider-Gold-Ginter system to incorporate the billing functionality of Wang. Gold discloses that the dedication service may allow senders “to purchase and send a dedication for a fee.” EX1005, [0009]. A POSITA would have understood that providing a service

to a user for a fee includes billing the user the amount of the fee. Moreover, Wang discloses *“bill[ing] a user...based on the purchase option.”* Given Wang’s disclosure of multiple purchase options for the identified content, as I discuss above for limitation [5.a], it would have been obvious to a POSITA that Wang’s system bills the user based on a selected purchase option because a bill would allow the provider of the fees to collect the fees. Weider in combination with Gold, Ginter, and Wang therefore renders limitation [5.b] obvious.

2. Claim 12: “The method of claim 11, further comprising: identifying, for the dedication, a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option; and billing a user associated with the first utterance based on the purchase option.”

177. Claim 12 is substantively identical to claim 5 and is rendered obvious for the same reasons.

E. Ground 5: Weider in view of Gold and Wang renders obvious claims 15-16

1. Claim 15

[15.a] “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: determine that a user is interacting with a media item; and”

178. Weider in combination with Gold and Wang renders this limitation obvious. As I discuss above, Gold’s jukebox Web server receives song search criteria from a sender and provides to the sender “a list of available songs meeting the criteria.” EX1005, [0007]. The sender “may elect to listen to a song from the

list to confirm that it is the song the sender wants,” before instructing the system to dedicate the song. *Id.*; *see id.*, [0031], [0032]; *see also id.*, [0048] (“Selecting a ‘Play Song’ link 430 causes the exemplary Web site to send the song to the sender’s web client so that the song can be played to the sender.”); FIG. 5 (excerpted):

Jukebox: Search Results

Your search returned 1 matches.

410	ID:	228	405
415	Artist:	David Bowie	
420	Song:	Fame	
425	Year:	75	
	Audio:	Play Song	Dedicate Song

430

| New Search | List All |

400

A POSITA would have recognized that a song is a “*media item*” and understood that the sender (user) “*interact[s] with the media item*” by playing the song on the sender device. The system “*determine[s] that [the] user is interacting with [the] media item*” when the user initiates playback of the song on the sender device (e.g., by selecting the “Play Song” link).

179. Wang also discloses “*determin[ing] that a user is interacting with a media item.*” Wang describes the “impulse purchase phenomenon” that allows “a music seller to take advantage of a potential impulse purchase resulting from a

listener hearing [the] song (for perhaps the first time) in a car or other location that is remote from normal retail locations.” EX1007, [0009]. To do so, Wang’s system identifies “songs and music from a relatively short sample and enabl[es] a user to interact with the service to immediately purchase a recording of the identified music remotely.” *Id.*, [0012]. Specifically, the user “capture[s] a sample of an audio stream to which the user is exposed” and “transmits the sample to a service provider” for identification using a “sampling device, such as an ordinary mobile (or cellular) phone.” *Id.*, [0019], [0020]. The audio stream may be a “short sample of music being played over a music playing device such as a radio” or “pre-recorded audio” that the user plays back on their device. *Id.*, [0012],[0020]; *see also id.*, [0056] (“[A] sample may first be recorded and then later transmitted to the system for identification.”).

180. A POSITA would have recognized that an audio stream/sample of music is a segment of a “**media item**” (e.g., a song) and would have understood that, by providing the real-time or pre-recorded sample for identification, Wang’s user is “**interacting with the media item.**” Wang’s signal identification service “**determin[es] that [the] user is interacting with [the] media item**” when the service receives the transmission and identifies the song (media item) being played.

181. In the combined Weider-Gold-Wang content dedication system, the user could select a song title to dedicate to a recipient from the song list of Gold

(e.g., by speaking the song title) or by sending a snippet of the song to the Weider-Gold server, which identifies the song as taught by Wang. A POSITA would have been motivated to take advantage of the “impulse purchase phenomenon” described in Wang by allowing a user listening to a song to immediately purchase the song for themselves or dedicate the song to another user rather than searching separately for the song as in Gold. This is a simple substitution of one known element (Gold’s search-based content identification techniques) for another (Wang’s signal processing-based content identification techniques) to obtain predictable results (identifying a song to dedicate to a recipient). Weider in combination with Gold and Wang therefore renders limitation [15.a] obvious.

[15.b] “identify, based on the natural language processing and the interaction of the user with the media item, the media item as the content to dedicate to the recipient.”

182. I note that the term “natural language processing” does not appear elsewhere in this claim or in independent claim 1. For purposes of this analysis, I interpret the term “natural language processing” as referring to the interpretation of the first utterance by the conversational language processor, which I discuss above for limitation [1.b(iv)].

183. Weider in combination with Gold and Wang renders this limitation obvious. As I discuss above for limitation [15.a], Gold’s sender interacts with a media item (song) by initiating playback of the media item on the sender device.

EX1005, [0007] (the sender “may elect to listen to a song from the list to confirm that it is the song the sender wants” to dedicate to the recipient). And as I discuss for limitations [1.b(ii)]-[1.b(iv)], Weider discloses using natural language processing techniques to interpret user utterances, including the first utterance. EX1004, [0093] (Weider’s system is an “interactive natural language processing system”), [0130] (Weider provides a “complete speech-based natural language command, algorithm and response environment”). In the combined system, the user speaks the first utterance to select a song title, interacts with the song by playing the song on the sender device, and speaks a command to “dedicate this song.” The system “*identif[es]...the media item as the content to dedicate to the recipient*” based on Weider’s “*natural language processing*” of the first utterance and Gold’s “*interaction of the user with the media item*” (i.e., the user initiating playback of the song).

184. Moreover, as I discuss above for limitation [15.a], Wang’s signal identification service determines that a user is interacting with a media item (song) when the service receives a transmission of an audio sample and identifies the song (media item) being played. EX1007, [0012], [0020]. In the combined Weider-Gold-Wang system, the user interacts with the system by voice, provides an audio sample that Wang’s signal identification service identifies as the song (media item), and says “dedicate this song.” The system “*identiffies]...the media item as the content*

to dedicate to the recipient” based on the “*natural language processing*” techniques of Weider, which includes processing the first utterance, and Wang’s “*interaction of the user with the media item*” (i.e., the user providing the song for identification by Wang’s signal identification service).

185. A POSITA would have been motivated to make this combination to extend the “impulse purchase phenomenon” described in Wang by allowing a user not only to purchase an identified song but to send a dedication containing the identified song to another user. Wang discloses that a user can provide input to the “interactive voice response” system using voice commands and it would have been obvious to use the natural language processing techniques of Weider to interpret the input. EX1007, [0037]. In fact, Wang explicitly recognizes that such techniques may be used, noting that where the user interacts with the system using voice prompts, “voice recognition methodologies are employed to translate the user’s voice responses into usable system inputs.” *Id.*, [0084].

186. Moreover, Wang allows a user to “interact with other users” of the service, for example, by “instruct[ing] the service to send a recommendation for the identified song to a friend...or a clip of the song to the friend’s mobile telephone.” *Id.*, [0043]; *see also id.*, [0042]. It would have been obvious that the recommendation could include a voice message and take the form of a dedication, as in Gold. This is a simple substitution of one known element (Gold’s search-based

content identification techniques) for another (Wang's signal processing-based content identification techniques) to yield predictable results (identifying a song to dedicate to a recipient). Weider in combination with Gold and Wang therefore renders limitation [15.b] obvious.

2. Claim 16

[16.a] “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: determine a media item that a user is playing;”

187. Weider in combination with Gold and Wang renders this limitation obvious. As I discuss above for limitation [15.a], Gold's sender can “[s]elect a ‘Play Song’ link 430 caus[ing] the exemplary Web site to send the song to the sender’s web client so that the song can be played to the sender.” EX1005, [00048]; *see also id.*, [0007], [0031], FIG. 5. As I discuss above, a POSITA would have recognized that a song is a “***media item***” and understood that Gold’s system “***determine[s] a media item that [the] user is playing***” when the sender (user) initiates playback of the song on the sender device.

188. Wang also discloses “***determin[ing] a media item that a user is playing.***” As I discuss above for limitation [15.a], Wang’s user “captures a relatively short sample of music being played over a music playing device” and transmits the sample to the signal identification service, which “identifies the song to the user in real-time.” EX1007, [0012]. The audio stream may be a “short sample of music

being played over a music playing device such as a radio” or “pre-recorded audio” that the user plays back on their device. *Id.*, [0012], [0020]; *see also id.*, [0056] (“[A] sample may be first be recorded and then later transmitted to the system for identification.”). A POSITA would have understood that Wang’s service “*determine[s] a media item that the user is playing*” when it identifies a song playing on the “music playing device.” Weider in combination with Gold and Wang therefore renders limitation [16.a] obvious.

[16.b] “determine that the user provided the first utterance during the playing of the media item by the user; and”

189. Weider in combination with Gold and Wang renders this limitation obvious. Wang discloses “*playing of the media item by the user.*” For example, the user “captures a relatively short sample of music being played over a music playing device” and transmits the sample to the service by “playing the music into the handset of the telephone.” EX1007, [0012].

190. Wang also discloses that the user can “*provide[] [an] utterance during the playing of the media item by the user.*” For example, Wang’s user can interact with the service by providing “voice commands” that may be made “while listening to the music at any location.” EX1007, [0002], [0037]; *see also id.*, [0059] (“By providing for real-time identification of songs being played or performed anywhere,” Wang’s system provides the user “with the song and artist information

that they wish to know...as soon as they hear it, i.e., at the peak of their interest in the song.”).

191. In the combined Weider-Gold-Wang system, the user “***provide[s] the first utterance***” while the song (media item) is playing on the user device, e.g., contemporaneously with the user providing the audio sample of the song to the server, as taught in Wang. Allowing the user to provide the first utterance “***during the playing of the media item by the user***” would have been an obvious extension of the “impulse purchase phenomenon” described in Wang by allowing the user playing the song to share it with another user. *See, e.g., id.*, [0043] (“[T]he user may instruct the service to send a recommendation for the identified song to a friend via email or SMS or a clip of the song to the friend’s mobile telephone”). The system “***determines that the user provided the first utterance***” using the natural language processing techniques of Weider, as I discuss above for limitations [1.b(ii)]-[1.b(iv)]. This combination involves use of a known technique (Wang receiving and processing voice commands while music is playing) to improve a similar device (the combined content dedication system) in the same way (supporting impulse-based user action). Weider in combination with Gold and Wang therefore renders limitation [16.b] obvious.

[16.c] “identify, based on the natural language processing and the determination that the user provided the first utterance during the playing of the media item, the media item as the content to dedicate to the recipient.”

192. Weider in combination with Gold and Wang renders this limitation obvious for the reasons I discuss above for limitation [15.b]. Additionally, the combined system renders obvious ***“identify[ing]...based on the...determination that the user provided the first utterance during the playing of the media item, the media item as the content to dedicate to the recipient.”*** As discussed above, Wang’s user can provide “voice commands” “while listening to the music at any location.” EX1007, [0002]. For example, the user can “instruct the service to send a recommendation for the identified song to a friend via email or SMS or a clip of the song to the friend’s mobile telephone.” *Id.*, [0043]. In the combined system, these “voice commands” can include the ***“first utterance”*** as I explain above in limitation [16.b]. The system then dedicates the content based on this ***“first utterance”*** made in conjunction with the ***“second utterance”*** while the media item is playing. Weider in combination with Gold and Wang therefore renders limitation [16.c] obvious.

F. Ground 6: Kennewick in view of Gold and Weider renders obvious claims 1, 3, 6-8, 10, 13-14, and 19

1. Claim 1

[1.pre] “A system for providing a natural language content dedication service, comprising:”

193. To the extent that the preamble is limiting, Kennewick in combination with Gold and Weider renders it obvious. Kennewick discloses ***“a system for***

providing a natural language...service.” Kennewick describes a “mobile interactive natural language speech system” that provides a “speech-based natural language query, response, and command environment.” EX1008, [0011]-[0012]. Specifically, Kennewick’s system includes a “speech interface device” that “receives spoken natural language queries, commands and/or other utterances from a user,” processes the input, and “responds to the user with a natural language speech response.” *Id.*, [0014]. For example, “a user may say, ‘tune in my favorite radio station,’” or provide a natural language command when the user “wish[es] to listen to a particular piece of music.” *Id.*, [0027], [0018].

194. For each utterance, Kennewick “processes the input” to interpret the user command, invokes an agent to obtain information, and constructs a response to the user. *Id.*, [0014]; *see also id.*, [0028]; FIG. 5 (reproduced below, illustrating the interactive natural language speech system, which receives a user utterance via Speech Unit 128 and processes the input at Main Unit 98):

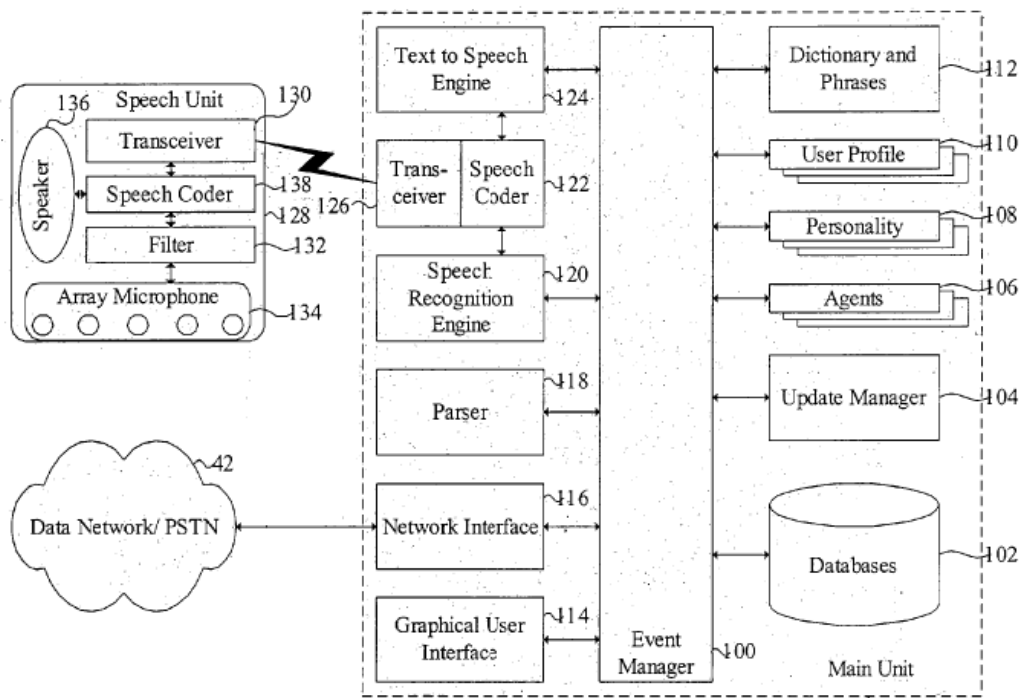


Figure 5. Speech Processing System Block Diagram

see also *id.*, [0095]-[0097], [0111], FIGS. 1-3, 6. The response is passed to a text-to-speech engine, which “creates the required utterances” that are “output by the speaker 136.” *Id.*, [0124].

195. A POSITA would have understood that Kennewick’s “mobile interactive natural language speech system” is a “***system for providing a natural language...service***” because a user provides input to the system via a natural language utterance (e.g., a spoken command or query) and receives, from the system, a natural language response. *Id.*, [0012]; see also *id.*, [0014]. The system provides a “service” to users by processing and responding to user utterances. For example, where the user utterance is a command to “tune in my favorite radio station,” the system can process the request and tune to that station. *Id.*, [0027]. In another

example, where the user provides a command to “listen to a particular piece of music,” a domain agent of the system “will make one or more queries...to locate a source for the music and retrieve it” and “activate a suitable player for the format of the music.” *Id.*, [0018].

196. To the extent Kennewick does not explicitly disclose that the system “*provid[es]...a content dedication service*,” the combination with Gold renders it obvious, as I discuss above for the preamble [1.pre] of Ground 1.

197. It would have been obvious to modify Kennewick to include Gold’s content dedication service and thereby provide a “*natural language content dedication service*.” Kennewick discloses a natural language service with music-related functionality, including processing and responding to user commands to “tune in my favorite radio station” or “listen to a particular piece of music.” EX1008, [0027], [0018]. A POSITA would have been motivated to add Gold’s content dedication to Kennewick’s service to enable users to share specific songs with others, since “song dedications can be an effective way to convey and share complex emotions and memories when it might not be possible, or desired, for the sender and recipient to communicate directly.” EX1005, [0006].

198. This is a combination of prior art elements according to known methods to yield predictable results. Kennewick’s system uses a flexible agent-based architecture to provide natural language services. EX1008, [0002] (“the invention

is a fully integrated environment allowing mobile users to ask natural language speech questions or give natural language commands in a wide range of domains”), [0017] (“The agents provide convenient and re-distributable packages or modules of functionality, typically for a specific domain....an agent may include everything that is needed to extend the functionality of the invention to a new domain.”). Gold’s content dedication functionality, i.e., sending and dedicating a selected song to a recipient, is one such type of domain. A POSITA would have found it obvious to use Gold’s functionality within Kennewick’s system, such as by using a domain agent like those taught by Kennewick to provide the content dedication service taught by Gold, and would have had a reasonable expectation for its success.

199. Alternatively, it would have been obvious to a POSITA to combine Gold and Kennewick for the reasons the Examiner identified during prosecution of the ’025 patent. EX1003, 187-98 (Final Office Action), 129-44 (Examiner’s Answer). Specifically, it would have been obvious to a POSITA to modify Gold by substituting Kennewick’s natural language command for Gold’s UI-based instruction to send a dedication to a recipient. For example, it would have been obvious to use Kennewick’s natural language service to process a natural language utterance (e.g., a content dedication request) and to route that utterance as input, e.g., to identify content, to Gold’s song dedication system. This is a simple substitution of one known element (Kennewick’s method for receiving and processing natural

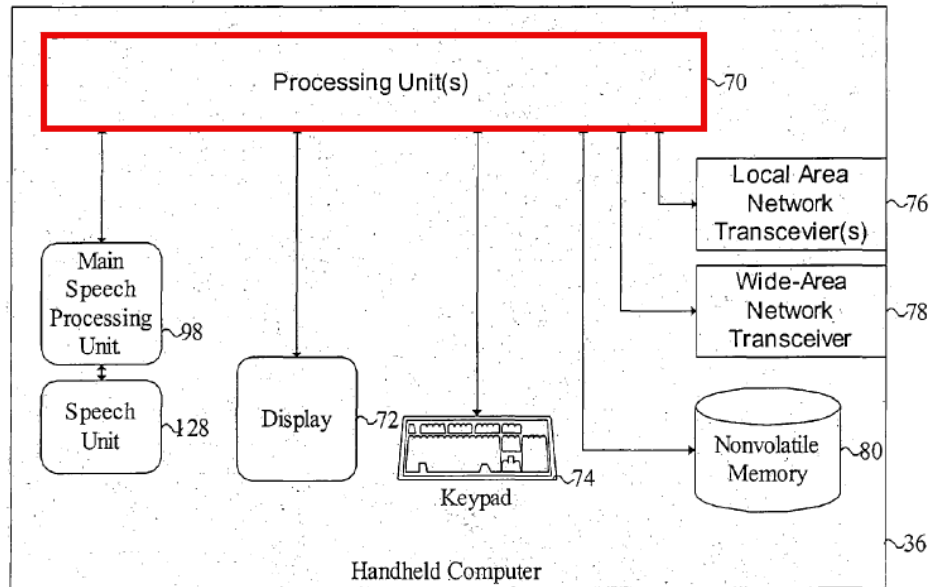
language utterances) for another (Gold’s song selection method). Moreover, as discussed above, Gold describes an embodiment in which the system is an “Interactive Voice Response (IVR) dedication system” with which the user interacts via telephone. EX1005, [0032]; *see also id.*, [0045] (The IVR features “allow[] the jukebox Web server to be accessed from a sender telephone.”). As the Examiner explained, “using Kennewick to process a natural language utterance, and route the utterance as input to the system of Gold, would have been recognized by those of ordinary skill in the art as resulting in an improved system that would allow selection of content based on natural language.” EX1003, 134-35. Kennewick in combination with Gold and Weider therefore renders preamble [1.pre] obvious.

200. I understand that the Examiner rejected preamble [1.pre] based on Kennewick and that the Applicant did not dispute the Examiner’s rationale. EX1003, 395-96 (Non-Final Office Action), 190 (Final Office Action), 145-65 (Appeal Brief).

[1.a]: “one or more processors; and;”

201. Kennewick in combination with Gold and Weider renders this limitation obvious. Kennewick’s mobile interactive natural language speech system “may include a stand alone or networked PC attached to a vehicle, a standalone or networked fixed computer in a home or office, a PDA, wireless phone, or other portable computer device, or other computer device or system.” EX1008, [0013];

see also *id.*, [0012], [0016]. Kennewick discloses that the handheld computer, which is illustrated in FIG. 3, includes “one or more processing units 70,” which “may include one or more central processing units.” *Id.*, [0111]; see also *id.*, FIG. 3:



Likewise, in the embodiment that uses a “fixed computer 44,” Kennewick discloses that the fixed computer 44 includes a “main speech processing unit 98” that “interface[s] to the fixed computer’s one or more processing units 84,” which “may include one or more central processing units.” *Id.*, [0115]; see also *id.*, [0095]-[0097], FIGS. 1-2, 4-6. A central processing unit is a processor. EX1024, 132 (defining a central processing unit as “the device that interprets and executes instructions” and a single-chip central processing unit as a “microprocessor”).

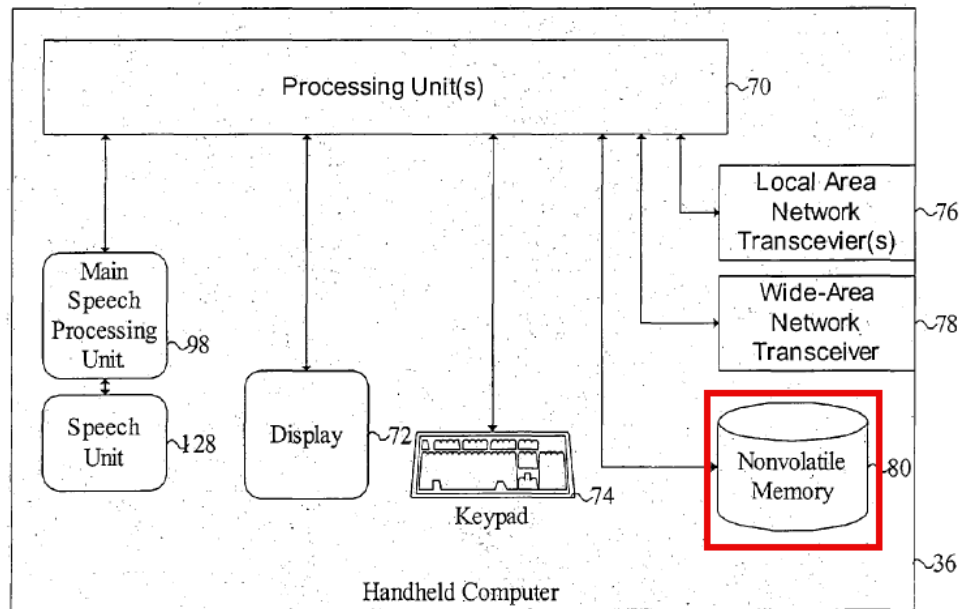
Therefore, Kennewick discloses that the system comprises “*one or more processors.*”

202. Gold also discloses “*one or more processors,*” as I discuss above for limitation [1.a] of Ground 1. Kennewick in combination with Gold and Weider therefore renders limitation [1.a] obvious.

203. I understand that the Examiner rejected limitation [1.a] based on Kennewick and that the Applicant did not dispute the Examiner’s rationale. EX1003, 191 (Final Office Action), 145-65 (Appeal Brief).

[1.b] “one or more non-transitory computer readable mediums storing executable instructions that when executed by the one or more processors cause the one or more processors to:”

204. Kennewick in combination with Gold and Weider renders this limitation obvious. Kennewick’s processing unit “uses one [or] more types of nonvolatile memory 80 for software and data storage. Suitable types of nonvolatile memory 80 include flash memory and hard disk drives.” EX1008, [0111]; *see also id.*, FIG. 3:



Likewise, in the embodiment that uses a “fixed computer 44,” Kennewick discloses nonvolatile memory 94 for “software and data storage.” *Id.*, [0114], [0115], FIG. 4. Nonvolatile memory is a type of non-transitory computer readable medium because it stores instructions and data readable by the processor. Kennewick therefore discloses “*one or more non-transitory computer readable mediums.*”

205. Kennewick also discloses that the computer readable medium “*stor[es] executable instructions.*” For example, Kennewick’s speech processing unit contains various software modules, including a “speech recognition engine,” a “parser,” and a “text to speech engine,” as well as “agents” that are “packages of executable code.” *Id.*, [0123], [0124], [0017]; *see also id.*, FIG. 4. A POSITA would have understood that software modules and packages of executable code comprise executable instructions (i.e., instructions that are executed by the one or more

processors) and are stored in the “*one or more non-transitory computer readable mediums*” (e.g., non-volatile memory 80). A POSITA also would have recognized that the instructions, “*when executed by the one or more processors cause the one or more processors to*” perform programmed operations because this is how computers conventionally operate. *See also id.*, [0012]-[0014], [0095]-[0097], FIGS. 1-2, 4-6.

206. Gold also discloses “*one or more non-transitory computer readable mediums storing executable instructions that when executed by the one or more processors cause the one or more processors to*” perform programmed operations, as I discuss above for limitation [1.b] of Ground 1. Kennewick in combination with Gold and Weider therefore renders limitation [1.b] obvious.

207. I understand that the Examiner rejected limitation [1.b] based on Kennewick and that the Applicant did not dispute the Examiner’s rationale. EX1003, 191 (Final Office Action), 145-65 (Appeal Brief).

[1.b(i)] “receive a first utterance that includes a natural language utterance;”

208. Kennewick in combination with Gold and Weider renders this limitation obvious. Kennewick’s interactive natural language speech system includes a “speech [unit] interface device that receives spoken natural language queries, commands and/or other utterances from a user.” EX1008, [0014]; *see also id.*, [0034] (“[T]he invention is intended to be able to accept most any natural

language question or command.”). For example, a user may provide a command to “listen to a particular piece of music” or instruct the system to “tune in my favorite radio station.” *Id.*, [0018], [0027]. A natural language utterance command to “listen to a particular piece of music,” for example, is “*a first utterance that includes a natural language utterance.*”

209. Kennewick also discloses “*receiving*” the “*first utterance.*” Kennewick’s speech unit interface device includes a speech unit 128, which includes “one or more microphones” to “receive the utterances of the user.” *Id.*, [0121]. “The speech received at the microphone 134 may be processed by filters 132 and passed to the speech coder 138 for encoding and compression.” *Id.*, *see also id.*, FIG. 5:

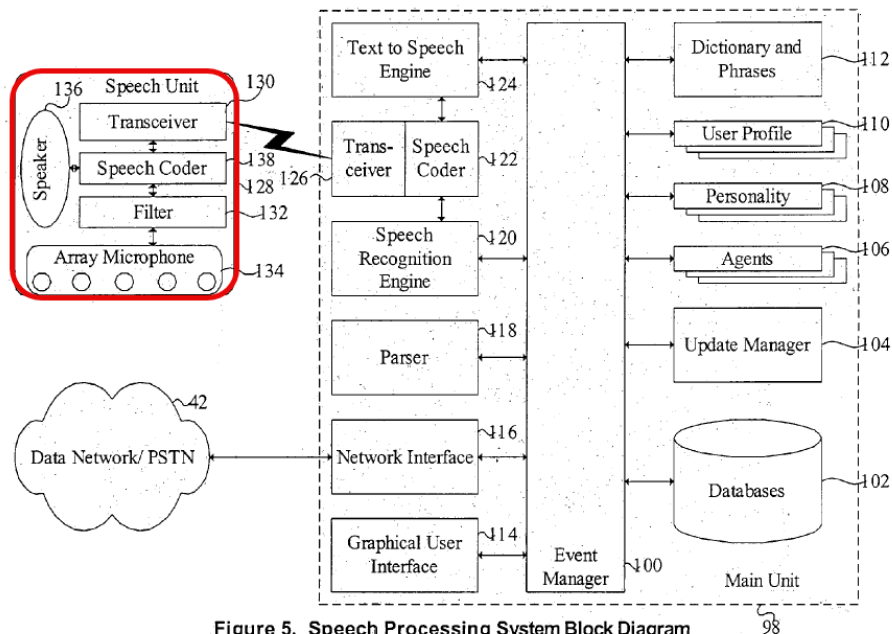


Figure 5. Speech Processing System Block Diagram

see also id., [0038] (“The system can use either one-dimensional or two-dimensional array microphones (or other devices) to receive human speech.”). Kennewick

“receive[s] [the] first utterance” captured through the speech unit 128. *See, e.g., id.*, [0028] (“For each user utterance...the system may perform multiple steps possibly including: 1) capturing the user’s question or query”); [0101] (“Virtually any service, involving the transfer of data or transmission of speech and video, can be supported through the natural language speech interface.”). Kennewick in combination with Gold and Weider therefore renders limitation [1.b(i)] obvious.

210. I understand that the Examiner rejected limitation [1.b(i)] based on Kennewick and that the Applicant did not dispute the Examiner’s rationale. EX1003, 191 (Final Office Action), 145-65 (Appeal Brief).

[1.b(ii)] “determine, based on processing of the first utterance by a speech recognition engine, one or more words or phrases of the first utterance;”

211. Kennewick in combination with Gold and Weider renders this limitation obvious. Speech input received at Kennewick’s natural language system, including the *“first utterance”* is “passed to the speech coder 122 for decoding and decompression.” *Id.*, [0123]. “The output of the speech coder 122 may be fed to the speech recognition engine 120,” which “recognizes words and phrases” by processing the decoded speech “using data in the dictionary and phrases module 112.” *Id.*, [0123], [0155]; *see also id.*, FIG. 5:

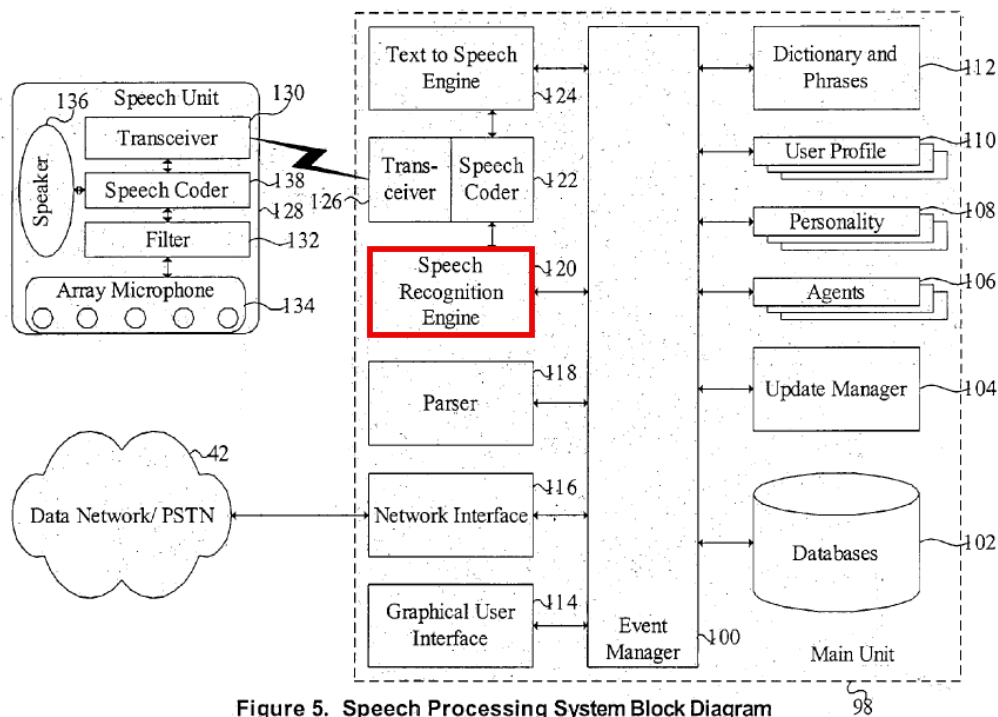


Figure 5. Speech Processing System Block Diagram

see also *id.*, [0028] (“For each user utterance, ... the system may perform multiple steps possibly including: 1) capturing the user’s question or query through accurate speech recognition operating in a variety of real-world environments.”), [0029]-[0034], [0041]. A POSITA would have recognized that Kennewick’s speech recognition engine meets the claimed “*speech recognition engine*” because it is “software or hardware that recognizes the words or phrases in the natural language utterance.” EX1021, 2.

212. Kennewick further discloses that the speech recognition engine “may determine the user’s identity by voice and name for each utterance” and that “[r]ecognized words and phrases are tagged with this identity in all further processing” such that “tags, added by the speech recognition engine 120 to each

utterance, may allow other components of the system to tie that utterance to the correct user and dialog.” EX1008, [0155].

213. A POSITA would have understood that, by recognizing words and phrases in the decoded speech, Kennewick’s speech recognition engine performs *“processing of [the] utterance”* to *“determine...one or more words or phrases of the [] utterance.”* See also *id.*, [0123] (“The decoded speech can be processed by the speech recognition engine 120.”). Such “processing” further includes tagging the utterance with a user identity. A POSITA also would have recognized that because these steps are performed for each utterance (*id.*, [0028], [0155]), the speech recognition engine *“determines one or more words or phrases of the first utterance”* (i.e., for the command to “listen to a particular piece of music”). Kennewick in combination with Gold and Weider therefore renders limitation [1.b(ii)] obvious.

214. I understand that the Examiner rejected limitation [1.b(ii)] based on Kennewick and that the Applicant did not dispute the Examiner’s rationale. EX1003, 191 (Final Office Action), 145-65 (Appeal Brief).

[1.b(iii)] “provide the one or more words or phrases as an input to a conversational language processor;”

215. Kennewick in combination with Gold and Weider renders this limitation obvious. “Once the words and phrases have been recognized by the speech recognition engine 120,” Kennewick’s speech recognition engine “pass[es]

these [words and phrases] to the parser 118 for interpretation.” *Id.*, [0160], [0155];
see also id., FIG. 5:

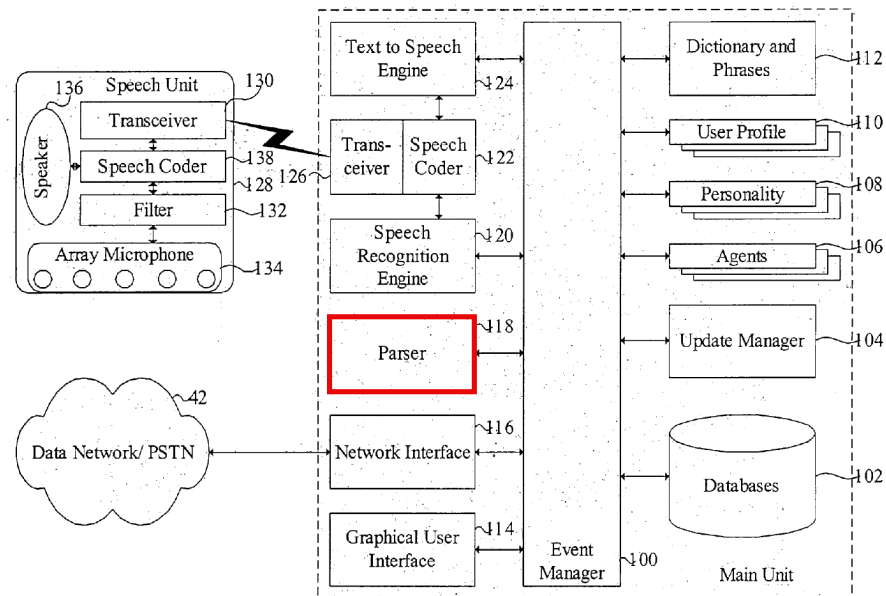


Figure 5. Speech Processing System Block Diagram

216. Specifically, Kennewick discloses that the speech recognition engine passes “tokens” corresponding to the recognized words and phrases to the parser, which “examines the tokens” and determines a context for the utterance “by applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, dialog history, and context stack contents.” *Id.*, [0160]. The context “may determine the domain and thereby the domain agent 156, if any, to be invoked,” and may correspond to the type of command or action to be performed. *Id.* The parser “may use a scoring system to determine the most likely context or domain for a user’s question or command” and “[b]ased on this scoring, the system may invoke the

correct agent.” *Id.*, [0161]; *see also id.*, [0127] (“Based on the keywords and command structure, the parser may invoke the required agent.”).

217. A POSITA would have recognized that Kennewick’s parser functions as a “*conversational language processor*” by determining an interpretation of the recognized words or phrases of the natural language utterance, e.g., by examining the tokens of the recognized words and phrases for context and “applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, dialog history, and context stack contents.” *Id.*, [0160]; *see also id.*, [0123] (“The recognized words and phrases may be processed by the parser 118, which transforms them into complete commands and questions using data supplied by the agents.”).

218. This is consistent with the disclosure of the ’025 patent, which describes that the conversational language processor “attempt[s] to determine an intent for the multi-modal interaction” by “determin[ing] a most likely context for the interaction from the preliminary interpretation of the utterance.” EX1001, 4:35-46. And like Kennewick’s parser, which “invoke[s] the correct agent” for an utterance based on the determined context, the conversational language processor of the ’025 patent “invoke[s] one or more agents 132c having capabilities for processing requests in a particular domain or application.” EX1008, [0161], EX1001, 13:8-20. A POSITA would have recognized, therefore, that Kennewick’s parser meets the claimed “*conversational language processor*.” Additionally, a

POSITA would have understood that by “passing” the recognized words and phrases to the parser (conversational language processor) for interpretation, the speech recognition engine “*provide[s] the one or more or phrases as an input to the conversational language processor.*” Kennewick in combination with Gold and Weider therefore renders limitation [1.b(iii)] obvious.

[1.b(iv)] “interpret the first utterance, at the conversational language processor, based on the one or more words or phrases;”

219. Kennewick in combination with Gold and Weider renders this limitation obvious for the reasons I discuss above for limitation [1.b(iii)]. A POSITA would have recognized that Kennewick’s parser (the “*conversational language processor*”) “*interprets the first utterance... based on the one or more words or phrases*” because, as discussed above, Kennewick’s speech recognition engine “passes” the recognized words and phrases “to the parser 118 for interpretation.” EX1004, [0187]; *see also id.*, [0028] (“For each user utterance... the system may perform multiple steps possibly including: ...parsing and interpreting the question or query.”). Specifically, a POSITA would have understood that the parser uses the “tokens” corresponding to the recognized words and phrases to determine a context for the utterance (e.g., a music context), thereby “interpreting” the first utterance. *See* [EX1008], [0160]. Kennewick in combination with Gold and Weider therefore renders limitation [1.b(iv)] obvious.

[1.b(v)] “identify, based on the interpretation of the first utterance, content to dedicate to a recipient;”

220. Kennewick in combination with Gold and Weider renders this limitation obvious. Kennewick discloses “***identify[ing], based on the interpretation of the first utterance, content.***” As I discuss above for the preamble [1.pre], Kennewick’s system retrieves a requested piece of music in response to a user command. EX1008, [0018]. A POSITA would have recognized that Kennewick’s natural language system first “identifies” a piece of music (content) to provide that music to the user and would have understood that this identification is based on the interpretation of the utterance by Kennewick’s parser (the “***conversational language processor***”) as I discuss above. *See, e.g.*, EX1008, [0018] (“a user may wish to listen to a particular piece of music...the domain agent will...locate a source for the music and retrieve it.”).

221. To the extent Kennewick does not explicitly disclose “***identify[ing]...content to dedicate to a recipient,***” the combination with Gold renders this obvious, as I discuss above for limitation [1.b(v)] of Ground 1. *See, e.g.*, EX1005, [0036] (“The sender reviews the song list and selects a song to dedicate to a recipient.”).

222. A POSITA would have found it obvious to integrate the content dedication functionality of Gold into the interactive natural language system of Kennewick. In the combined system, a user interacts with Kennewick’s natural

language speech interface to select a song title, causing the system to “*identify...content to dedicate to a recipient*” (i.e., the user-specified song). Kennewick discloses that the user can use the speech interface to select a song title. EX1008, [0018] (“a user may wish to listen to a particular piece of music”). A POSITA would have found it obvious for the user to use the natural language interface to command the system to dedicate the song as I discuss for the preamble [1.pre]. A POSITA would have been motivated to make this combination to further supplement the “wide range of domains” with which Kennewick’s speech interface may be used. *Id.*, [0002].

223. This combination involves application of a known technique (Gold’s content dedication functionality) to improve a similar device (Kennewick’s interactive natural language system) in the same way (enabling users to dedicate content).

224. Alternatively, as I discuss above for the preamble [1.pre], it would have been obvious to combine Gold and Kennewick for the reasons the Examiner identified during prosecution of the ’025 patent. EX1003, 187-98 (Final Office Action), 129-44 (Examiner’s Answer). Specifically, it would have been obvious to a POSITA to modify Gold by substituting Kennewick’s technique of interpreting content using natural language utterances for Gold’s UI-based technique for

identifying content. A POSITA would have been motivated to make this combination for the reasons I discuss above for the preamble [1.pre].

225. Kennewick in combination with Gold and Weider therefore renders limitation [1.b(v)] obvious.

226. I understand that the Examiner rejected a limitation similar to issued limitation [1.b(v)] based on Kennewick and Gold and that the Board affirmed the rejection. EX1003, 133-35 (Examiner's Answer), 79-82 (Decision on Appeal).

[1.b(vi)] “initiate, based on the identified content, a dedication to the recipient;”

227. Kennewick in combination with Gold and Weider renders this limitation obvious. Kennewick discloses “*initiat[ing]*” an action “*based on the identified content.*” For example, as I discuss above, when a user provides a natural language command to “listen to a particular piece of music,” Kennewick's system “locate[s] a source for the music and retrieve[s] it” and “activate[s] a suitable player for the format of the music.” EX1008, [0018]. A POSITA would have recognized that each of these actions is “*initiate[d] based on the identified content.*” For example, a POSITA would have understood that the piece of music that is “retrieved” is the music (content) that is identified in limitation [1.b(v)] above.

228. Gold discloses “*initiat[ing]...a dedication to the recipient,*” as I discuss above for limitation [1.b(vi)] of Ground 1. *See, e.g.*, EX1005, [0036]-[0038], FIG. 2.

229. A POSITA would have been motivated to combine Kennewick and Gold to provide a natural language content dedication service for the reasons I discuss for the preamble [1.pre], and a POSITA also would have found it obvious to “*initiate...the dedication to the recipient*” as taught by Gold. For example, a POSITA would have found it obvious to use Kennewick’s speech interface to provide the user with instructions for recording the dedication voice message. *See, e.g.,* EX1008, [0172] (an agent can “[a]sk the user for more information, typically through the speech interface”), [0124] (an agent may output speech to a user). A POSITA would have been motivated to combine the references to create “a complete speech-based information query, retrieval, presentation and command environment.” *Id.*, Abstract, [0011]. This combination involves application of a known technique (Gold’s content dedication functionality) to improve a similar device (Kennewick’s interactive natural language system) in the same way (enabling users to dedicate content). Kennewick in combination with Gold and Weider therefore renders limitation [1.b(vi)] obvious.

230. I understand that the Examiner rejected limitation [1.b(vi)] based on Gold and that the Board affirmed the rejection. EX1003, 191 (Final Office Action), 83 (Decision on Appeal).

[1.b(vii)] “receive a second utterance to be associated with the dedication;”

231. Kennewick in combination with Gold and Weider renders this limitation obvious. Weider discloses that the user may provide “speech annotations” for objects stored by a server. EX1004, [0120], [0121] (“the annotated objects may be stored on a server side”). For example, the user can provide a speech annotation for a content object that is “stored along with the object[.]” *Id.*, [0120]. The object can be “digital media.” *Id.*, claim 23; *see also id.*, [0120]. A POSITA would have recognized a speech annotation as an utterance, as I discuss above for [1.b(vii)] of Ground 1.

232. Gold, in turn, discloses “***receiving***” an utterance “***to be associated with the dedication,***” as I discuss above for limitation [1.b(vii)] of Ground 1. *See, e.g.*, EX1005, Abstract, [0038]-[0039].

233. A POSITA would have found it obvious to use Kennewick’s natural language speech system to receive the dedication voice message (the second utterance) from the user. In the combination, the user uses Kennewick’s speech interface to provide a “***first utterance***” that the system uses to “***identify content to dedicate to a recipient,***” as I discuss for limitation [1.b(v)]. A POSITA would have found it obvious for the user to similarly provide, and the system to “***receive a second utterance to be associated with the dedication.***” Kennewick discloses receiving multiple utterances from a user, including requests for music-related

functionality. EX1008, [0018], [0155], [0037]. Weider discloses receiving voice annotations (utterances) for digital media objects. EX1004, [0120], [0121]. Gold discloses receiving a voice message from the dedication sender. EX1005, [0031], claim 6. As I discuss above for Ground 1, Weider discloses a system that enables users to annotate a digital media object by voice. EX1004, [0120], [0127], claims 19, 23. Weider does not limit what a user may say in these speech annotations. EX1004, [0120]. One example of a speech annotation is described in Gold—a dedication voice message. It would have been obvious to a POSITA to try Gold’s dedication message as Weider’s speech annotation because “song dedications can be an effective way to convey and share complex emotions and memories.” EX1005, [0006]. In the combination, the “*second utterance*” is Weider’s speech annotation, and specifically Gold’s dedication voice message. The combination furthers Kennewick’s objective to create “a complete natural language speech-based query and response environment.” EX1008, [0006]. Kennewick in combination with Gold and Weider therefore renders limitation [1.b(vii)] obvious.

234. I understand that the Examiner rejected limitation [1.b(vii)] based on Gold and that the Applicant did not dispute the Examiner’s rationale. EX1003, 190-191 (Final Office Action), 145-65 (Appeal Brief).

[1.b(viii)] “determine, based on a processing of the second utterance by the speech recognition engine, one or more words or phrases of the second utterance;”

235. Kennewick in combination with Gold and Weider renders this limitation obvious. As I discuss above for limitation [1.b(vii)], Kennewick in combination with Gold and Weider renders obvious ***“receiving a second utterance to be associated with the dedication”*** (the speech annotation dedication message). And as I discuss for limitation [1.b(ii)], the combination renders obvious ***“determin[ing], based on processing of the first utterance by a speech recognition engine, one or more words or phrases of the first utterance.”*** A POSITA would have found it obvious to perform the same speech recognition processing described above for the second utterance to allow the speech recognition engine to ***“determine...one or more words or phrases”*** of the dedication message because Kennewick discloses performing speech recognition processing on “each user utterance.” EX1008, [0028] (“For each user utterance, ...the system may perform multiple steps possibly including: 1) capturing the user’s question or query through accurate speech recognition operating in a variety of real-world environments.”); *see also id.*, [0029], [0030][0034], [0041], [0123].

236. Moreover, as I discuss for limitation [1.b(vii)], Weider’s speech annotation of an object is an utterance. EX1004, [0120] (The system is “configured to allow speech annotations of objects stored thereon.”), [0127] (disclosing “verbal annotations” of objects). To annotate the objects, Weider discloses that the system

“transcribe[s] the speech annotations to textual annotations and store[s] the textual annotations with the object.” *Id.*, [0120], [0127] (similar), claim 19 (similar). A POSITA would have understood that, by transcribing the speech annotation (utterance) to a textual annotation, Weider’s speech recognition engine “*process[es] the second utterance*” to “*determine...one or more words or phrases of the second utterance*” (speech annotation dedication message). Kennewick in combination with Gold and Weider therefore renders limitation [1.b(viii)] obvious.

237. I understand that the Examiner rejected limitation [1.b(viii)] based on Kennewick and that the Applicant did not dispute the Examiner’s rationale. EX1003, 194 (Final Office Action), 145-65 (Appeal Brief).

[1.b(ix)] “provide the one or more words or phrases of the second utterance as textual annotations within metadata of the content; and”

238. Kennewick in combination with Gold and Weider renders this limitation obvious. As I discuss above for limitation [1.b(ix)] of Ground 1, Weider discloses storing “*textual annotations within metadata of the content.*” For example, Weider describes that the system “may be configured to allow speech annotations of objects stored thereon,” where the objects include content, such as “digital media.” EX1004, [0120], [0127] (similar), claim 23. The system “transcribe[s] the speech annotations to textual annotations and store[s] the textual annotations with the object.” *Id.*, [0120]; *see also id.*, [0123] (“the textual annotations and/or textual descriptions that are associated with the objects may be

stored as metadata”). As I discuss above for limitation [1.b(ix)] of Ground 1, the ’025 patent confirms that Weider’s disclosure of storing annotations as metadata meets *“provid[ing] the one or more words or phrases...as textual annotations within metadata of the content.”* See, e.g., EX1001, 29:56-64.

239. Kennewick in combination with Gold and Weider renders obvious *“receiv[ing] a second utterance to be associated with the dedication,”* as I discuss for limitation [1.b(vii)]. Given this combination, a POSITA would have found it obvious to transcribe the second utterance (speech annotation dedication message) and store the words or phrases of that utterance as *“textual annotations within metadata of the content”* in view of the combined teachings of the references. A POSITA would have been motivated to make this combination for the reasons I provide above for limitation [1.b(vii)] and because storing the annotation as metadata (as disclosed by Weider) would have provided an efficient way of conveying the dedication text of Gold to the recipient along with the content (i.e., sending the content and the metadata together) in a single message instead of multiple messages that would result in consumption of additional network resources like bandwidth. Kennewick in combination with Gold and Weider therefore discloses limitation [1.b(ix)].

[1.b(x)] “send information to enable the recipient to access the content and the second utterance.”

240. Kennewick in combination with Gold and Weider renders this limitation obvious. In the combined system, the “***content***” is the song identified by the system as I discuss for limitation [1.b(v)]. The “***second utterance***” is the speech annotation dedication message as I discuss for [1.b(vii)], the words or phrases of which are “***provide[d] as textual annotations within metadata of the content,***” as I discuss for limitation [1.b(ix)]. Gold discloses “***send[ing] information to enable the recipient to access the content and the second utterance***” (i.e., the song and the dedication voice message), as I discuss above for limitation [1.b(x)] of Ground 1.

241. In the combination, the one or more words or phrases of the second utterance are “***provid[ed] as textual annotations within metadata of the content.***” A POSITA would have found it obvious to include the metadata of the content with the content, so that the recipient accessing the content also accesses the metadata. As I discuss for limitation [1.b(ix)], Weider discloses storing textual annotations within metadata of content. EX1004, [0123] (“the textual annotations and/or textual descriptions that are associated with the objects may be stored as metadata”). Weider also discloses that metadata may be included within the content. *Id.*, claim 28 (“wherein the object includes metadata”). Therefore, a POSITA would have found it obvious to provide the metadata of the content to the recipient along with

the content by including the metadata within the content or by otherwise sending it at the same time.

242. A POSITA would have found it obvious for the combined system of Kennewick, Gold, and Weider to inform the recipient of the dedication and “*enable the recipient to access the content and the second utterance*” as taught by Gold. Informing the recipient is the primary purpose of the dedication. EX1005, [0006] (“Song dedications can be an effective way to convey and share complex emotions and memories when it might not be possible, or desired, for the sender and recipient to communicate directly.”). In the combined system, the user can provide Kennewick’s natural language system with contact information for the recipient, e.g., by voice, and the combined system sends a dedication to the recipient using the contact information, e.g., by sending an email or phone call or otherwise sending a message using the appropriate contact information. This combination is a combination of prior art elements according to known methods (Kennewick collecting user input via the speech interface and Gold prompting a user to provide recipient contact information) to yield predictable results (using voice input to collect recipient contact information). Kennewick in combination with Gold and Weider therefore renders limitation [1.b(x)] obvious.

243. I understand that the Examiner rejected limitation [1.b(x)] based on Gold and that the Applicant did not dispute the Examiner's rationale. EX1003, 191 (Final Office Action), 145-65 (Appeal Brief).

2. Claim 3: "The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to annotate the content with the second utterance."

244. Claim 3 is substantively identical to limitation [1.b(ix)] and is rendered obvious for the same reasons.

3. Claim 6

[6.a] "The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: send, to the recipient, an electronic link to access the content as the information to enable the recipient to access the content; and"

245. Kennewick in combination with Gold and Weider renders this limitation obvious, as I discuss above for limitation [6.a] of Ground 1. *See also* EX1005, [0042], Abstract, claim 8.

246. I understand that the Examiner rejected limitation [6.a] based on Gold and that the Applicant did not dispute the Examiner's rationale. EX1003, 192 (Final Office Action), 145-65 (Appeal Brief).

[6.b] "send, to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance."

247. Kennewick in combination with Gold and Weider renders this limitation obvious, as I discuss above for limitation [6.b] of Ground 1. *See also* EX1005, [0039].

248. I understand that the Examiner rejected limitation [6.b] based on Gold and that the Applicant did not dispute the Examiner's rationale. EX1003, 192 (Final Office Action), 145-65 (Appeal Brief).

4. Claim 7: "The system of claim 1, wherein the content includes one or more of a song or a video."

249. Kennewick in combination with Gold and Weider renders claim 7 obvious for the same reasons as I discuss above in Ground 1. *See also* EX1005, [0007].

250. I understand that the Examiner rejected claim 7 based on Gold and that the Applicant did not dispute the Examiner's rationale. EX1003, 192 (Final Office Action), 145-65 (Appeal Brief).

5. Claim 8

251. Kennewick in combination with Gold and Weider renders claim 8 obvious for similar reasons as claim 1. As I discuss above for the preamble [1.pre] and limitations [1.a]-[1.b], Kennewick in combination with Gold and Weider renders obvious "[a] system for providing a natural language content dedication service," where the system comprises "*one or more processors*" and "*one or more non-transitory computer readable mediums storing executable instructions that when executed by the one or more processors cause the one or more processors to*" perform programmed operations.

252. A POSITA would have recognized that the combined Kennewick-Gold-Weider system performs “*a method for providing a natural language content dedication service*” for the same reasons. For example, Kennewick is titled “Mobile Systems and Methods for Responding to Natural Language Speech Utterance” and is directed to “[m]obile systems and methods” that use “a complete speech-based information query, retrieval, presentation, and command environment.” EX1008, Abstract; *see also id.*, [0011]. Similarly, as I discussed above for claim 8 of Ground 1, Gold is directed to “a system and method for using the Internet or electronic media employing computerized database approaches for tying together songs and personal voice messages to send song dedications.” EX1005, [0007]; *see also id.*, claim 1 (“A method of sending a song dedication from a sender to a recipient.”). A POSITA would have understood that the systems of Kennewick and Gold, which I discuss above, perform the described methods, including “*a method for providing a natural language content dedication service.*” The remaining limitations of claim 8 are substantively identical to claim 1. Kennewick in combination with Gold and Weider therefore renders claim 8 obvious.

6. Claim 10: “The method of claim 8, further comprising annotating the content with the second utterance.”

253. Claim 10 is substantively identical to claim 3 and is rendered obvious for the same reasons.

7. Claim 13: “The method of claim 8, further comprising: sending, to the recipient, an electronic link to access the content as the information to enable the recipient to access the content; and sending, to the recipient, the second utterance in an encoded audio format as the information to enable the recipient to access the second utterance.”

254. Claim 13 is substantively identical to claim 6 and is rendered obvious for the same reasons.

8. Claim 14: “The method of claim 8, wherein the content includes one or more of a song or a video.”

255. Claim 14 is substantively identical to claim 7 and is rendered obvious for the same reasons.

9. Claim 19: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide the second utterance within the content.”

256. Kennewick in combination with Gold and Weider renders this limitation obvious for the same reasons I discuss above for claim 19 of Ground 1. *See also* EX1005, [0040].

257. I understand that the Examiner rejected claim 19 based on Gold and that the Applicant did not dispute the Examiner’s rationale. EX1003, 192 (Final Office Action), 145-65 (Appeal Brief).

G. Ground 7: Kennewick in view of Gold, Weider, and Martinez renders obvious claims 2, 9, 17, and 18

1. Claim 2: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide an image as one or more tags within metadata of the content.”

258. Kennewick in combination with Gold, Weider, and Martinez renders claim 2 obvious. As I discuss above for limitation [1.b(ix)], Weider discloses *“provid[ing] the one or more words or phrases of the second utterance as textual annotations within metadata of the content.”* And as I discuss for claim 2 of Ground 2, Martinez further discloses *“provid[ing] an image as one or more tags within metadata of the content.”*

259. It would have been obvious to a POSITA to modify the combined Kennewick-Gold-Weider system to include the image metadata of Martinez. For example, a POSITA would have found it obvious for the content dedication service to provide an image (e.g., album art) with a dedicated song to allow the recipient to more easily identify the song, the artist, and/or the album on which the song is included. *See, e.g.* EX1006, [0038] (a media item may include “a jpeg image media item portion”); *id.*, [0041] (“[T]he media item portion may be any type of identifiable portion or component of media item,” such as “a graphical image”). Providing the image as a *“tag[] within metadata of the content”* would have provided an efficient way of conveying the image to the recipient along with the content (i.e., sending the

content and the image metadata together), e.g., by keeping the image and content packaged together, improving playback and display consistency, and simplifying the system architecture.

260. As I discuss above in Ground 2, this is a combination of prior art elements according to known methods to yield predictable results. Both Weider and Martinez acknowledge that metadata may include many different types of data. *See, e.g.,* EX1004, [0123] (“One of ordinary skill in the art will readily appreciate that metadata may include various types of information.”); EX1006, [0047] (“[A] metadata subset may comprise any number and combination of information or data about a media item portion”). One such data type disclosed in Martinez includes image tags. EX1006, [0045] (metadata may include a “graphical image such as a jpeg”). Combining Martinez’s image metadata with the Kennewick-Gold-Weider system that permits different metadata types yields the predictable result of the metadata including the image tags of Martinez. Kennewick in combination with Gold, Weider, and Martinez therefore renders claim 2 obvious.

2. Claim 9: “The method of claim 8, further comprising providing an image as one or more tags within metadata of the content.”

261. Claim 9 is substantively identical to claim 2 and is rendered obvious for the same reasons.

3. Claim 17: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide an electronic link associated with the content as one or more tags within metadata of the content.”

262. Kennewick in combination with Gold, Weider, and Martinez renders claim 17 obvious. As I discuss above for claim 2, Kennewick in combination with Weider and Martinez discloses “*providing [data] as one or more tags within metadata of the content.*” Moreover, as I discuss above for claim 17 of Ground 2, Martinez discloses that the data includes “*an electronic link associated with the content.*”

263. It would have been obvious to a POSITA to modify the combined Kennewick-Gold-Weider system to “*provide an electronic link associated with the content as one or more tags within metadata of the content,*” as disclosed by Martinez. For example, where the metadata includes an image (e.g., album art), a POSITA would have recognized that providing a pointer or an electronic link to the image within the metadata would have reduced the metadata file size and enabled faster load and transfer times than if the metadata contained the image data itself. A POSITA would have been motivated to make this combination for the reasons I provide above for claim 2. Kennewick in combination with Weider and Martinez therefore renders claim 17 obvious.

4. Claim 18: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to provide a voice tag as one or more tags within metadata of the content.”

264. Kennewick in combination with Gold, Weider, and Martinez renders claim 18 obvious. As I discuss above for claim 2, Kennewick in combination with Weider and Martinez discloses “*providing [data] as one or more tags within metadata of the content.*” Moreover, as I discuss above for claim 18 of Ground 2, Martinez discloses that the data includes a “*voice tag.*” A POSITA would have been motivated to make this combination for the reasons I provide above for claim 2. Kennewick in combination with Weider and Martinez therefore renders claim 18 obvious.

H. Ground 8: Kennewick in view of Gold, Weider, and Ginter renders obvious claims 4 and 11

1. Claim 4: “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to purchase the content from a content provider to enable the recipient to access the content.”

265. Kennewick in combination with Gold, Weider, and Ginter renders claim 4 obvious. As I discuss for the preamble [1.pre], Gold’s content dedication system includes a “database of songs for use in the dedication.” EX1005, [0007]. And as I discuss for claim 4 of Ground 3, Ginter discloses “*purchas[ing] the content from a content provider.*”

266. It would have been obvious to modify the combined Kennewick-Gold-Weider system by incorporating Ginter's disclosed method for purchasing content items from content creators. For example, Ginter discloses that the purchased content includes music. *See, e.g.*, EX1009, 1:58-61 ("The owners and marketers of this content include...recording companies"), 56:48-51 ("[A] musical performing group may act as both content creator 102 and distributor 106 by creating and distributing its own musical recordings."), 59:28-33 ("Information content 304 could be...sound such as a musical performance"), 141:8-14 ("[C]ontent may include:...music"). Gold discloses that the content dedication service includes a song database. EX1005, [0007]. In the combined system, Gold's service functions as a distributor, and the content creator of Ginter is the content provider. *See, e.g.*, EX1009, 3:48-59 ("[T]he present invention...can help ensure that content providers will be paid for use of distributed information."), 2:27-59 ("Commercial content providers are concerned with ensuring proper compensation for the use of their electronic information."); *see also id.*, FIG. 80. It would have been obvious for Gold to populate the song database by purchasing the content from a content creator (as disclosed by Ginter) because this is how music is conventionally acquired, *i.e.*, users and services buy the right to use songs from the rights holders. *See, e.g.*, EX1016, [0010] ("[T]o obtain endorsement [by the music industry], companies must charge users for access to the songs."), [0051] (A user "engage[s] in an on-line commerce

transaction in which the user pays for access rights to [a] particular media item.”); EX1009, 3:19-21 (“VDE allows the owners and distributors of electronic digital information to reliably bill for, and securely control, audit, and budget the use of, electronic information.”). A POSITA would have been motivated to make this combination because, for example, Ginter discloses storing purchased content in a content repository for distribution to users. EX1009, 309: 25-38, 311:44-51, 314:1-2, 315:59-316:15, FIG. 78.

267. In the combination, the Kennewick-Gold-Weider-Ginter content dedication system purchases content from content providers (e.g., artists or record companies) and functions as a distributor by sending (distributing) information “*to enable the recipient to access the content*,” as I discuss above for limitation [1.b(x)]. The use of Ginter’s disclosed method of purchasing content to build the song database of Gold is a combination of prior art elements according to known methods – Gold’s distribution of database content to users (dedication recipients) and Ginter’s method of acquiring content via purchase from content creators – to yield predictable results – a song dedication system populated with content purchased from content creators. The result of the combination is predictable because conventional business relationships include sales transactions. A POSITA would have reasonably expected the combination to be successful for the same reasons. Kennewick in combination with Gold, Weider, and Ginter therefore renders claim 4 obvious.

2. Claim 11: “The method of claim 8, further comprising purchasing the content from a content provider to enable the recipient to access the content.”

268. Claim 11 is substantively identical to claim 4 and is rendered obvious for the same reasons.

I. Ground 9: Kennewick in view of Gold, Weider, Ginter, and Wang renders obvious claims 5 and 12

1. Claim 5

[5.a] “The system of claim 4, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: identify, for the dedication, a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option; and”

269. Kennewick in combination with Gold, Weider, Ginter, and Wang renders this limitation obvious. As I discuss above for the preamble [1.pre] and limitation [1.b(iii)], Kennewick’s interactive natural language speech system uses a flexible agent-based architecture to provide natural language services. EX1008, [0017] (“[D]omain specific behavior and information are organized into agents” that “receive, process and respond to user questions, queries and commands.”). Kennewick further discloses that the system “may provide license management capability allowing the sale of agents by third parties to one or more users on a one time or subscription basis.” EX1008, [0019]. Gold describes providing a specific service – sending and dedicating a selected song to a recipient. A POSITA would have found it obvious to use Gold’s functionality within Kennewick’s system, such

as by using a domain agent like those taught by Kennewick to provide the content dedication service taught by Gold. In the combined system, the domain agents performing the content dedication service are provided on a subscription basis (a “*paid subscription purchase option*”) as disclosed by Kennewick. Kennewick in combination with Gold, Weider, Ginter, and Wang therefore renders limitation [5.a] obvious.

270. Wang also discloses “*identify[ing], for the dedication, a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option,*” as I discuss above for limitation [5.a] of Ground 4.

271. A POSITA would have found it obvious to modify the combined Kennewick-Gold-Weider-Ginter system to incorporate the disclosed purchase options taught by Wang. First, it would have been obvious to implement Wang’s “*buy-to-own*” purchase option for the dedication within the combined system to allow a recipient of a dedication to download the dedicated song and replay the song on command. This would have been an obvious way to implement the embodiment of Gold in which the system provides the dedication (including the dedicated song) in a physical computer file (e.g., as an attachment to the email notification). EX1005, [0008]. Second, Gold discloses that “senders may use the invention to purchase and send a dedication for a fee,” and it would have been obvious for the

fee to take the form of a paid subscription for use of Gold's content dedication service, as taught by Wang or a "buy-to-own" fee, e.g., for each dedication the user sends. *Id.*, [009]. The result of the combination is predictable because both "***buy-to-own***" and "***paid subscription***" purchase models for content and use of services were well known. A POSITA would have reasonably expected the combination to be successful for the same reason.

272. Moreover, as I discuss above for claim 4, Kennewick in combination with Gold, Weider, and Ginter renders obvious "***purchas[ing] the content from a content provider to enable the recipient to access the content.***" Where the identified purchase option is "***buy-to-own***," it would have been obvious to a POSITA that a portion of the user payment would be passed to the content provider as part of the payment that the distributor (Gold's service) makes to the content provider (Ginter's content creator) to purchase the content. This is consistent with the disclosure of the '025 patent, which recites that the amount charged to the user "encompasses the cost for purchasing the content from the content provider 680 plus a service charge for dedicating the content." EX1001, 30:54-58. Alternatively, where the identified purchase option is a "***paid subscription purchase option***," it would have been obvious for the combined system to use a subscription service to recuperate the cost of building the song database, as I discuss above for claim 4.

273. A POSITA would have recognized that the various disclosed purchase options would have enabled the operator of the combined system to generate profit for providing a service and would have found it obvious to incorporate purchase options into the combined Kennewick-Gold-Weider-Ginter system because this is how services typically operate. *See, e.g.,* EX1016, [0010] (“[T]o obtain endorsement [by the music industry], companies must charge users for access to the songs.”), [0051] (A user “engage[s] in an on-line commerce transaction in which the user pays for access rights to [a] particular media item.”). This combination includes use of a known technique (Weider’s digital download or subscription-based monetization model) to improve a similar product (a digital content system) in the same way (enabling revenue generation). Kennewick in combination with Gold, Weider, Ginter, and Wang therefore renders limitation [5.a] obvious.

274. I understand that the Examiner rejected limitation [5.a] based on Kennewick and that the Applicant did not dispute the Examiner’s rationale. EX1003, 195 (Final Office Action), 86 (Decision on Appeal).

[5.b] “bill a user associated with the first utterance based on the purchase option.”

275. Kennewick in combination with Gold, Weider, Ginter, and Wang renders this limitation obvious for the same reasons I discuss above for limitation [5.b] of Ground 4.

276. I understand that the Examiner rejected limitation [5.b] based on Ginter and that the Applicant did not dispute the Examiner's rationale. EX1003, 195 (Final Office Action), 86 (Decision on Appeal).

2. Claim 12: "The method of claim 11, further comprising: identifying, for the dedication, a purchase option that includes one or more of a buy-to-own purchase option, a pay-to-play purchase option, or a paid subscription purchase option; and billing a user associated with the first utterance based on the purchase option."

277. Claim 12 is substantively identical to claim 5 and is rendered obvious for the same reasons.

J. Ground 10: Kennewick in view of Gold, Weider, and Wang renders obvious claims 15-16

1. Claim 15

[15.a] "The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: determine that a user is interacting with a media item; and"

278. Kennewick in combination with Gold, Weider, and Wang renders this limitation obvious, as I discuss above for limitation [15.a] of Ground 5.

[15.b] "identify, based on the natural language processing and the interaction of the user with the media item, the media item as the content to dedicate to the recipient."

279. As I discuss above for limitation [15.b] of Ground 5, I interpret the term "natural language processing" as referring to the interpretation of the first utterance by the conversational language processor, which I discuss above for limitation [1.b(iv)].

280. Kennewick in combination with Gold, Weider, and Wang renders this limitation obvious. As I discuss above for limitation [15.a], Gold’s sender “*interact[s] with a media item*” (song) by initiating playback of the media item on the sender device. EX1005, [0007] (the sender “may elect to listen to a song from the list to confirm that it is the song the sender wants” to dedicate to the recipient). And as I discuss for limitations [1.b(ii)]-[1.b(iv)], Kennewick discloses using natural language processing techniques to interpret user utterances, including the first utterance. EX1008, [0012] (Kennewick’s system is an “interactive natural language speech system”), [0011] (“Through this integrated approach, a speech-based natural language query, response and command environment is created.”). In the combined system, the user speaks the first utterance to select a song title, interacts with the song by playing the song on the sender device, and speaks a command to “dedicate this song.” The system “*identif[es]...the media item as the content to dedicate to the recipient*” based on Kennewick’s “*natural language processing*” of the first utterance and Gold’s “*interaction of the user with the media item*” (i.e., the user initiating playback of the song).

281. Moreover, as I discuss above for limitation [15.a], Wang’s signal identification service “*determine[s] that a user is interacting with a media item*” (song) when the service receives a transmission of an audio sample and identifies the song (media item) being played. EX1007, [0012], [0020]. In the combined

Kennewick-Gold-Weider-Wang system, the user interacts with the system by voice, provides an audio sample that Wang's signal identification service identifies as the song (media item), and says "dedicate this song." The system "*identifies]...the media item as the content to dedicate to the recipient*" based on the "*natural language processing*" techniques of Kennewick, which includes processing the first utterance, and Wang's "*interaction of the user with the media item*" (i.e., the user providing the song for identification by Wang's signal identification service).

282. As I discuss above in Ground 5, a POSITA would have been motivated to make this combination to extend the "impulse purchase phenomenon" described in Wang by allowing a user not only to purchase an identified song but to send a dedication containing the identified song to another user. Wang discloses that a user can provide input to the interactive voice response system using voice commands (EX1007, [0037]), and it would have been obvious to use the natural language processing techniques of Kennewick to interpret the input. In fact, Wang explicitly recognizes that such techniques may be used, noting that where the user interacts with the system using voice prompts, "voice recognition methodologies are employed to translate the user's voice responses into usable system inputs." *Id.*, [0084].

283. Moreover, Wang allows a user to "interact with other users" of the service, for example, by "instruct[ing] the service to send a recommendation for the

identified song to a friend...or a clip of the song to the friend's mobile telephone.” *Id.*, [0043]; *see also id.*, [0042]. It would have been obvious that the recommendation includes a voice message and takes the form of a dedication, as in Gold. This is a simple substitution of one known element (Gold's search-based content identification techniques) for another (Wang's signal processing-based content identification techniques) to obtain predictable results (identifying a song to dedicate to a recipient). Kennewick in combination with Gold, Weider, and Wang therefore renders limitation [15.b] obvious.

2. Claim 16

[16.a] “The system of claim 1, wherein the executable instructions, when executed by the one or more processors, cause the one or more processors to: determine a media item that a user is playing;”

284. Kennewick in combination with Gold, Weider, and Wang renders this limitation obvious for the same reasons I discuss above for limitation [16.a] of Ground 5.

[16.b] “determine that the user provided the first utterance during the playing of the media item by the user; and”

285. Kennewick in combination with Gold, Weider, and Wang renders this limitation obvious. As I discuss above for limitation [16.b] of Ground 5, Wang discloses “*provid[ing] [an] utterance during the playing of the media item by the user.*” In the combined Kennewick-Gold-Weider-Wang system, the user “*provide[s] the first utterance*” while the song (“*media item*”) is playing on the user

device, e.g., contemporaneously with the user providing the audio sample of the song to the server, as taught in Wang. Allowing the user to “*provide[] the first utterance during playing of the media item*” would have been an obvious extension of the “impulse purchase phenomenon” described in Wang by allowing the user playing the song to share it with another user. *See, e.g.,* EX1007, [0043] (“[T]he user may instruct the service to send a recommendation for the identified song to a friend via email or SMS or a clip of the song to the friend’s mobile telephone”). The system “*determines that the user provided the first utterance*” using the natural language processing techniques of Kennewick, as I discuss above for limitations [1.b(ii)]-[1.b(iv)]. This combination involves use of a known technique (Wang receiving and processing voice commands while music is playing) to improve a similar device (the combined content dedication system) in the same way (supporting impulse-based user action). Kennewick in combination with Gold, Weider, and Wang therefore renders limitation [16.b] obvious.

[16.c] “*identify, based on the natural language processing and the determination that the user provided the first utterance during the playing of the media item, the media item as the content to dedicate to the recipient.*”

286. Kennewick in combination with Gold, Weider, and Wang renders this limitation obvious for the reasons I discuss above for limitation [15.b]. Additionally, the combined system renders obvious “*identify[ing]...based on the...determination that the user provided the first utterance during the playing of the media item, the*

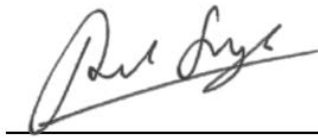
media item as the content to dedicate to the recipient.” As discussed above, Wang’s user can interact provide “voice commands” “while listening to the music at any location.” EX1007, [0002]. For example, the user can “instruct the service to send a recommendation for the identified song to a friend via email or SMS or a clip of the song to the friend’s mobile telephone.” *Id.*, [0043]. In the combined system, these “voice commands” can include the “*first utterance*” as I explain above in limitation [16.b]. The system then dedicates the content based on this “*first utterance*” made in conjunction with the “*second utterance*” while the media item is playing. Kennewick in combination with Gold, Weider, and Wang therefore renders limitation [16.c] obvious.

XI. CONCLUSION

287. For the reasons I have presented above, it is my opinion that the references serving as the bases for the grounds of rejection in the Petition render claims 1-19 of the challenged patent unpatentable.

I hereby 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.

Dated: July 30, 2025

A handwritten signature in black ink, appearing to read "Pad Smyth", written over a horizontal line.

Padhraic Smyth