

**IN THE UNITED STATES DISTRICT COURT  
FOR THE WESTERN DISTRICT OF TEXAS  
MIDLAND/ODESSA DIVISION**

Dialect, LLC,

Plaintiff,

v.

Meta Platforms, Inc.,

Defendant.

Case No. 7:25-cv-00060-DC-DTG

**DEFENDANT'S PRELIMINARY INVALIDITY CONTENTIONS**

September 17, 2025

## TABLE OF CONTENTS

	<u>Page</u>
I. PRELIMINARY STATEMENT AND RESERVATION OF RIGHTS .....	1
II. BACKGROUND KNOWLEDGE OF A PERSON OF ORDINARY SKILL AND THE STATE OF THE ART IN SPEECH RECOGNITION AND NATURAL LANGUAGE UNDERSTANDING.....	4
A. Recognizing Words and Phrase in Spoken Natural Language Utterances.....	6
B. Using Prior Probabilities, e.g., Hidden Markov Models, in Speech Recognition.....	12
C. Using Multimodal Input with Speech Recognition Technology .....	14
D. Accessing Information with Domain Agents .....	16
E. Using User Personalization Models with Speech Recognition Technology.....	17
III. IDENTIFICATION OF PRIOR ART .....	20
A. Background Prior Art.....	21
B. Charts Setting Forth Where in the Prior Art References Each Element of the Asserted Claim(s) Are Found .....	25
1. '209 Patent .....	27
2. '006 Patent .....	27
3. '825 Patent .....	28
4. '607 Patent .....	29
5. '039 Patent .....	30
C. Obviousness.....	30
1. '209 Patent .....	33
2. '006 Patent .....	39
3. '825 Patent .....	44
4. '607 Patent .....	50
5. '039 Patent .....	60
IV. INVALIDITY UNDER 35 U.S.C. § 112.....	68
A. Section 112 ¶ 2: Indefiniteness.....	68

B.	Section 112 ¶ 1: Lack of Written Description .....	71
V.	PATENT-ELIGIBILITY UNDER 35 U.S.C. § 101 .....	73
A.	'209 Patent.....	76
B.	'006 Patent.....	76
C.	'607 Patent.....	77
D.	'039 Patent.....	77
E.	'825 Patent.....	78

Defendant Meta Platforms, Inc. (“Defendant”) submits these Preliminary Invalidation Contentions pursuant to the Scheduling Order (Dkt. 78) and the Court’s Standing Order Governing Proceedings (OGP) 4.4 – Patent Cases (“OGP”). These disclosures, together with the accompanying claim charts, set forth Defendant’s current and Preliminary Invalidation Contentions with respect to the asserted claims identified by the infringement contentions served by Plaintiff Dialect, LLC (“Dialect” or “Plaintiff”) on July 16, 2025 (the “Infringement Contentions”) for U.S. Patent Nos. 7,398,209 (the “’209 patent”), 8,015,006 (the “’006 patent”), 8,447,607 (the “’607 patent”), 9,263,039 (the “’039 patent”), and 9,734,825 (the “’825 patent”) (collectively, the “Asserted Patents”), as modified by the email from Justin Constant dated August 4, 2025. At present, the claims asserted by Dialect (“Asserted Claims”) are as follows:

**’209 patent:** claims 1, 4, 6, 7, 8, and 9;

**’006 patent:** claims 5–6;

**’607 patent:** claims 12–13;

**’039 patent:** claims 13–15, 17, and 18;

**’825 patent:** claims 5–8.

## **I. PRELIMINARY STATEMENT AND RESERVATION OF RIGHTS**

As set forth below, each Asserted Claim is invalid pursuant to 35 U.S.C. §§ 101, 102, 103, and/or 112. Defendant reserves the right to amend and/or supplement these Preliminary Invalidation Contentions, including for example based on information learned through further discovery or investigation. *See, e.g.*, OGP at 13 n.13. Defendant further reserves the right to amend and/or supplement these Preliminary Invalidation Contentions should Dialect provide additional information, amend or supplement its Infringement Contentions in any way, or otherwise change, supplement or purport to clarify its Infringement Contentions.

The following disclosures, including the accompanying claim charts, were prepared prior

to the Court's ruling on claim construction and prior to claim-construction disclosures by the parties. Defendant reserves the right to amend, supplement, or provide further invalidity contentions following the Court's *Markman* ruling, including providing final invalidity contentions according to the Scheduling Order (Dkt. 78).

Until the Court provides its claim construction rulings, Defendant's Preliminary Invalidity Contentions are based on the claim constructions apparently underlying Dialect's Infringement Contentions, to the extent that those Infringement Contentions are comprehensible and Dialect's constructions are discernible. These Preliminary Invalidity Contentions are not intended to, and may not, reflect Defendant's positions as to the proper construction of the Asserted Claims. To the extent that these Preliminary Invalidity Contentions reflect an interpretation consistent with any construction adopted by Dialect, no inference is intended, and none should be drawn, that Defendant agrees with Plaintiff's claim constructions, and Defendant expressly reserves its right to contest such constructions. Further, no inference is intended, and none should be drawn, that any Asserted Claim satisfies 35 U.S.C. § 112; and in fact, the Asserted Claims do not for the reasons set forth below.

Prior art not included in these contentions, whether or not currently known to Defendant, may become relevant, or its significance may change, depending on the claim construction positions that the parties assert and/or that the Court adopts, or on Dialect's theories of infringement, which are still largely inscrutable due to the deficiencies in Dialect's Infringement Contentions. Defendant's ongoing investigations may also uncover additional prior art or evidence related thereto, including through responses to third-party subpoenas and depositions. For example, for any given prior art system, Defendant expects that additional documentation and other evidence relating to these systems will be discovered, and Defendant reserves the right to

rely on such evidence to demonstrate the invalidity of the Asserted Claims. Defendant accordingly reserves the right to modify these Preliminary Invalidity Contentions, including without limitation by adding or withdrawing prior art and/or modifying any of the referenced claim charts in light of new developments or analysis, including without limitation claim-construction proceedings, any amended or supplemental infringement contentions by Dialect, any party admissions or other admissions concerning the scope of the claims or teachings of the prior art, or any positions taken by Dialect in this or related litigation, patent office review, or other proceeding, or to avoid unfair prejudice from Dialect's failure to timely comply with its disclosure obligations.

The identity of each item of prior art relied upon in these contentions is listed below and in the attached claim charts, which are incorporated herein by reference. These Preliminary Invalidity Contentions provide exemplary citations showing where the cited references disclose and/or render obvious each element of the Asserted Claims. All of the references cited and discussed herein may also be relied upon to show the state of the art in the relevant time frames. Defendant expressly reserves the right to rely on uncited portions of the prior art references, other documents, and expert testimony to provide context and/or to aid in understanding the cited portions of the references and to establish the invalidity of the Asserted Claims. Where a reference contains multiple disclosures of a given element, Defendant has generally cited only a subset of those disclosures, and Defendant reserves the right to rely on others. Similarly, where Defendant cites a particular figure in a reference, the citation should be understood to encompass the caption and description of the figure and any text relating to or discussing the figure. Conversely, where Defendant cites text referring to or relating to a figure, the citation should be understood to include the figure as well, and all other text referring or relating to that figure.

For purposes of these Preliminary Invalidity Contentions only, Defendant has applied the

purported earliest priority date as reflected on the face of the Asserted Patents. But the burden is on Dialect to demonstrate it is entitled to its asserted priority date and it has not met its burden. Defendant's analysis and review is thus ongoing and Defendant reserves the right to challenge the asserted priority dates and rely upon any intervening prior art in the event of a successful challenge. Moreover, to date, Defendant has not deposed the named inventors of the patents, Dialect has failed to produce all documents evidencing conception and reduction to practice (and requisite diligence) of the claimed inventions of the Asserted Patents, and Dialect has not substantiated any priority date earlier than the earliest filing date reflected on the face of the Asserted Patents, either based on conception and reduction to practice documents, or otherwise. In the event that the Asserted Patents are found to be entitled to an earlier date of invention than the purported filing date of the Asserted Patents, Defendant reserves the right to supplement these Preliminary Invalidity Contentions with additional prior art on this basis.

## **II. BACKGROUND KNOWLEDGE OF A PERSON OF ORDINARY SKILL AND THE STATE OF THE ART IN SPEECH RECOGNITION AND NATURAL LANGUAGE UNDERSTANDING**

By approximately mid-2002, the earliest priority date of any of the Asserted Patents,<sup>1</sup> the art of computer-based speech recognition and natural language understanding was rich and deep, having matured over several decades by drawing on sophisticated, technological contributions from research in the subsidiary and related fields of signal processing, physics (acoustics), pattern recognition, communication and information theory, linguistics, physiology, computer science,

---

<sup>1</sup> The Court's OGP requires a plaintiff to serve preliminary infringement contentions that, among other things, "identify the priority date (*i.e.*, the earliest date of invention) for each asserted claim." OGP at 2. In its July 16, 2025 Initial Infringement Contentions, Dialect only identified, as the earliest date of invention for each Asserted Patent, the filing date of the earliest application listed in the priority chain on the face of each Asserted Patent. As noted above in the Preliminary Statement and Reservation of Rights, Defendant reserves the right to amend and/or supplement its invalidity contentions in the event that Dialect identifies a different or earlier priority date.

and psychology.

In 1952, in one early example of automated speech recognition, researchers at Bell Laboratories built a system for isolated digit recognition for a single speaker, which “relied heavily on measuring spectral resonances during the vowel region of each digit.” Lawrence Rabiner & Biing-Hwang Huang, *FUNDAMENTALS OF SPEECH RECOGNITION* at 6 (1993); *see also* K. H. Davis et al., *Automatic Recognition of Spoken Digits*, *J. Acoustic Soc. Am.*, Vol. 24, No. 6, pp. 637–642 (1952). In 1956, researchers at RCA Laboratories developed a system to “recognize 10 distinct syllables of a single talker,” which also “relied on spectral measurements (as provided by an analog filter bank) primarily during vowel regions.” Lawrence Rabiner & Biing-Hwang Huang, *FUNDAMENTALS OF SPEECH RECOGNITION* at 6 (1993); *see also* H. F. Olson & H. Belar, *Phonetic Typewriter*, *J. Acoustic Soc. Am.*, Vol. 28, No. 6, pp. 1072–1081 (1956). In the 1960s, “three key research projects were initiated that . . . had major implications on the research and development of speech recognition” for the following 30 years. Lawrence Rabiner & Biing-Hwang Huang, *FUNDAMENTALS OF SPEECH RECOGNITION* at 7 (1993). Another “achievement of note in the 1960s was the pioneering research of [Raj] Reddy in the field of continuous speech recognition by dynamic tracking of phonemes. Reddy’s research eventually spawned a long and highly successful speech-recognition research program at Carnegie Mellon University (to which Reddy moved in the late 1960s) which, to this day, remains a world leader in continuous-speech-recognition systems.” *Id.* In 1973, Dr. Reddy and others at CMU developed and demonstrated HEARSAY-I, a computer-based speech recognition system that used acoustic-phonetics, prosodics, syntax, semantics, pragmatics, and a lexicon to recognize spoken English utterances and understand and respond to them. *See* Raj Reddy, *CMU Hearsay 1973 (Short Version)*, <https://www.youtube.com/watch?v=c70QlwttNvg&t=6s>. In 1975, James Baker—who along with

his wife, Janet Baker, went on to found Dragon Systems—published an overview of a “Dragon” continuous-speech recognition system that relied on probabilistic functions of Markov chain (a/k/a hidden Markov models, or HMMs). James K. Baker, *The DRAGON System—An Overview*, IEEE Trans. Acoustics, Speech, Signal Proc., Vol. ASSP-23, No. 1, pp. 24–29 (Feb. 1975). The technology of automated speech recognition continued evolving and advancing over the ensuing decades with input and impetus from the likes of AT&T, IBM, DARPA, SRI International, MIT, Bolt Beranek and Newman (BBN), and others. *See generally* Lawrence Rabiner & Biing-Hwang Huang, FUNDAMENTALS OF SPEECH RECOGNITION at 6–9 (1993). For example, “the 1980s was a decade in which a major impetus was given to large vocabulary, continuous-speech-recognition systems by the Defense Advanced Research Projects Agency (DARPA) community, which sponsored a large research program aimed at achieving high word accuracy for a 1000-word, continuous-speech-recognition, database management task.” *Id.* at 8.

As demonstrated in these invalidity contentions, the Asserted Patents added nothing significant or technological to this rich body of work that had already matured for approximately 50 years before the earliest priority date of any of the Asserted Patents.

#### **A. Recognizing Words and Phrase in Spoken Natural Language Utterances**

By no later than the late 1990s, the state of the art in computer-based speech recognition technology had advanced to the point that off-the-shelf speech recognition software was commercially available to consumers and could be run with very effective results on entry-level computers and with little to no training. This included, for example, continuous, large-vocabulary speech recognition systems such as Dragon NaturallySpeaking and IBM ViaVoice, which allowed users to perform command and control of PC applications as well as continuous dictation (e.g., in word-processing applications), and which would recognize words and phrases in users’ spoken utterances with high accuracy and essentially in real time. In other words, performing speech

recognition by recognizing words and phrases in spoken natural language utterances was well known—indeed, that was the whole point—and effective techniques for doing so were well understood. It was also well understood to perform speech recognition using lexicons to recognize words and phrases, such as lexicons of sub-word units, whole word units, and even multiple-word units. These lexicons, or dictionaries, mapped the basic speech units (e.g., sub-word units) detected in the acoustic speech signal to words and phrases in the language being spoken. For example, the canonical 1993 text FUNDAMENTALS OF SPEECH RECOGNITION included the below block diagram illustrating a general model of a “continuous speech-recognition system based on subword speech units”:

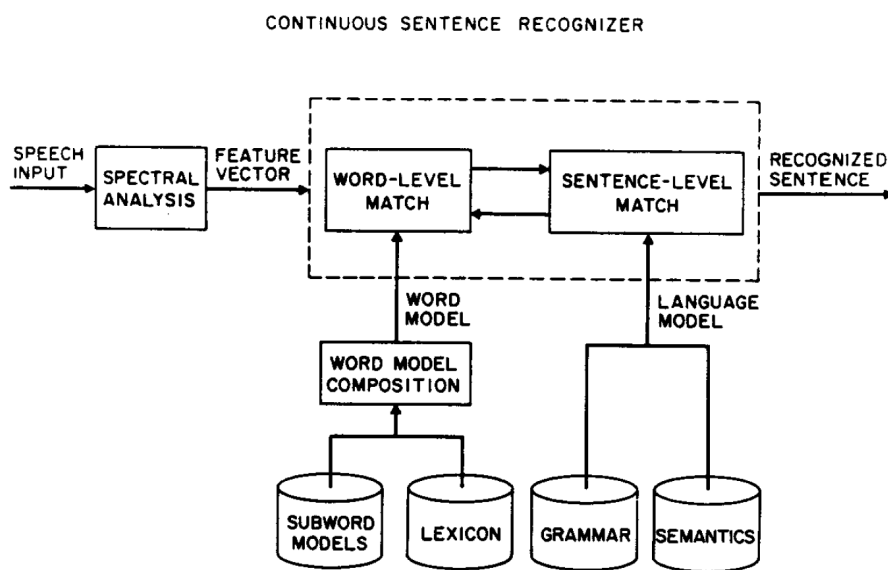


Figure 8.7 Overall block diagram of subword unit based continuous speech recognizer.

Lawrence Rabiner & Biing-Hwang Huang, FUNDAMENTALS OF SPEECH RECOGNITION at 450-451 (1993). As the authors explained, the lexicon is used in a speech recognition system “to describe the composition of words” in terms of their constituent subword units. *Id.* at 435; *see also id.* at 39 (explaining that, in the example of isolated digit recognition, “[t]he sound lexicon describes the pronunciations of digits in terms of the basic sounds of English,” and providing an example of

such a lexicon in Table 2.3). In simple terms, the acoustic signal representing the speech input would be processed to detect and identify subword units (e.g., phonemes), and the lexicon served as an index of probable word matches that likely correspond to that series of subword units detected in the speech signal. *See id.* at 450–451 (“The first step in the processing is spectral analysis to derive the feature vector used to characterize the spectral properties of the speech input. . . . The second step in the recognizer is a combined word-level/sentence-level match. The way this is accomplished is as follows. Using the set of subword HMMs and the word lexicon, a set of word models (HMMs) is created by concatenating each of the subword unit HMMs as specified in the word lexicon.”). The below table, also from FUNDAMENTALS OF SPEECH RECOGNITION, depicts a simple lexicon that maps a series of phonelike units (PLUs)—one type of subword unit—to words in the English language:

**TABLE 8.2. Typical word pronunciations (word lexicon) based on context-independent PLUs.**

Word	Number of phones	Transcription				
a	1	ax				
above	4	ax	b	ah	v	
bad	3	b	ae	d		
carry	4	k	ae	r	iy	
define	5	d	iy	f	ay	n
end	3	eh	n	d		
gone	3	g	ao	n		
hours	4	aw	w	axr	z	

*Id.* at 438. Indeed, FUNDAMENTALS OF SPEECH RECOGNITION devoted an entire chapter to discussing different techniques for “designing and implementing a word lexicon (which provides a mapping between words and subword units).” *Id.* at 6.

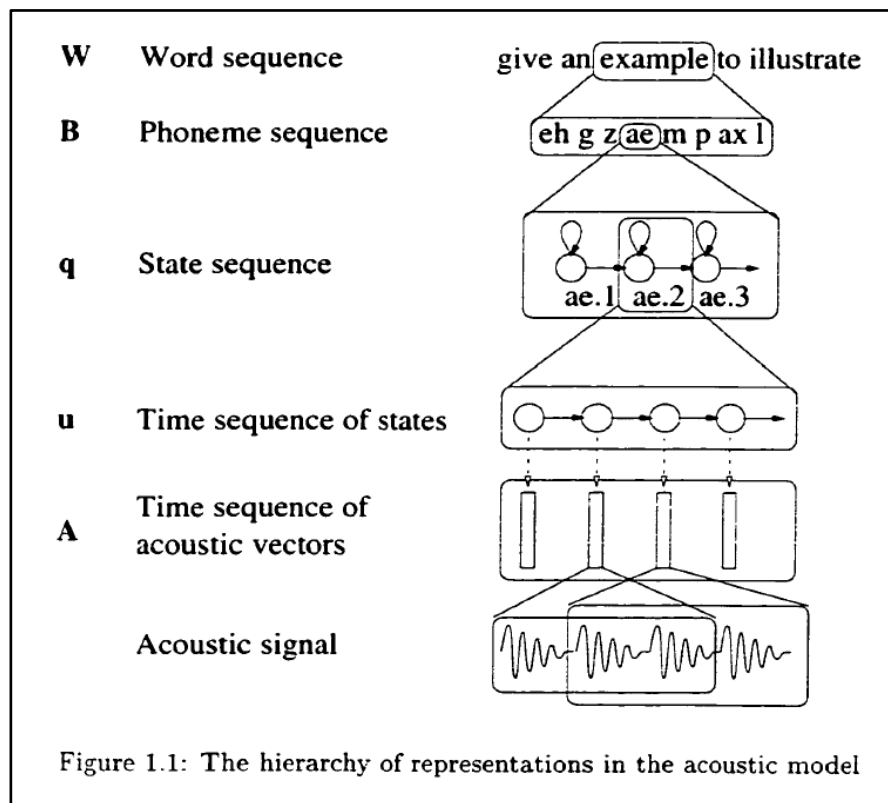
The authors further explained that use of subword units (and a corresponding lexicon that

maps each recognizable word in the vocabulary to a series of subword units) was advantageous for continuous, large-vocabulary speech recognition because it reduced the number of patterns that the system would need to look for as well as the amount of data that would be needed to train the speech recognition system. *See id.* at 436–437 (“Hence some more efficient speech representation is required for such large vocabulary systems. This is essentially the reason we use subword speech units.”). For example, rather than training and storing a whole-word pattern for each of the tens of thousands of possible words in a large vocabulary, a speech recognition system could use as few as 50 PLUs (shown below in Table 8.1), translating the PLUs detected and identified in the acoustic signal into words in the vocabulary using a word lexicon:

**TABLE 8.1. Set of basic PLUs for speech.**

Number	Symbol	Word	Number	Symbol	Word
1	h#	silence	26	k	kick
2	aa	father	27	l	led
3	ae	bat	28	m	mom
4	ah	butt	29	n	no
5	ao	bought	30	ng	sing
6	aw	bough	31	ow	boat
7	ax	again	32	oy	boy
8	axr	diner	33	p	pop
9	ay	bite	34	r	red
10	b	bob	35	s	sis
11	ch	church	36	sh	shoe
12	d	dad	37	t	rot
13	dh	they	38	th	thief
14	eh	bet	49	uh	book
15	el	bottle	40	uw	boot
16	en	button	41	v	very
17	er	bird	42	w	wet
18	ey	bait	43	y	yet
19	f	fief	44	z	zoo
20	g	gag	45	zh	measure
21	hh	hag	46	dx	butter
22	ih	bit	47	nx	center
23	ix	roses			
24	iy	beat			
25	jh	judge			

*Id.* at 438. This paradigm for speech recognition was well understood and widely known. For example, Saraçlar explained in a 2001 thesis: “Large vocabulary continuous speech recognition (LVCSR) systems typically represent words as concatenations of smaller subword units. Most such systems use a phonemic representation of a word, often called a baseform, which is usually obtained by looking up the word in a pronunciation dictionary which contains single pronunciations for most words. This provides a transformation from a word sequence **W** to a baseform phoneme sequence **B**. . . . For instance, a baseform for the word **example** is the phoneme sequence **eh g z ae m p ax l**, with  $J = 8$  phonemes. Another possible baseform for this word may be **eh g z ae m p e l**, with  $J = 7$  phonemes.” Murat Saraçlar, *Pronunciation Modeling for Conversational Speech Recognition* at 8–9 (Jan. 2001). Saraçlar illustrated this paradigm of using a dictionary to map words in the vocabulary to a sequence of subword units (in this illustration, to phonemes) with the below diagram:



*Id.* at 8.

Use of subword units was also known to have the advantage of reducing the number of samples needed for training a large-vocabulary speech recognition system. For example, while only a few thousand training samples might be needed for digit recognition (i.e., to recognize 0-9 plus “oh” in spoken utterances) if using whole-word models, expanding the vocabulary to 1,000 recognizable words would very quickly require hundreds of millions of training samples, in order to account for all the linguistic variability in the pronunciation of words: “For word vocabularies like the digits, we know that each digit can be preceded and followed by every other digit; hence for an 11-digit vocabulary (zero to nine plus oh), there are exactly 121 phonetic contexts (some of which are essentially identical). Thus with a training set of several thousand digit strings, it is both realistic and practical to see every digit in every phonetic context several times. Now consider a vocabulary of 1000 words with an average of 100 phonetic contexts for both the beginning and end of each word. To see each word in each phonetic context exactly once requires  $100 \times 1000 \times 100 = 10$  million carefully designed sentences. To see each combination 10 times requires 100 million such sentences. Clearly the recording and processing of such h[u]mo[ng]ous amounts of speech data is both impractical and unthinkable.” Lawrence Rabiner & Biing-Hwang Huang, *FUNDAMENTALS OF SPEECH RECOGNITION* at 436 (1993).

The authors of *FUNDAMENTALS OF SPEECH RECOGNITION* also explained that it may be advantageous to model whole phrases—which it referred to as “function phrases”—as standalone units due to the fact that they are “generally unstressed in speech,” “poorly articulated in continuous speech,” and “highly variable in pronunciation depending on context.” *Id.* at 475–476. Examples include phrases such as “in the” and “what is.” *Id.* The authors explained that “[e]xperience shows small but consistent improvements in recognition performance when function

word dependent units are added to the standard set of subword units.” *Id.*

### **B. Using Prior Probabilities, e.g., Hidden Markov Models, in Speech Recognition**

By no later than the late 1970s, it was also well understood to perform speech recognition using probabilistic techniques. For example, James Baker’s 1975 overview of the Dragon continuous-speech recognition system explained that “[t]he general model which is used throughout the DRAGON system is that of a probabilistic function of a Markov process.” James K. Baker, *The DRAGON System—An Overview*, IEEE Trans. Acoustics, Speech, Signal Proc., Vol. ASSP-23, No. 1, pp. 24–29 at 24 (Feb. 1975). By the 1990s, probabilistic techniques for speech recognition were routine; indeed, by that time, “[t]he standard approach to large vocabulary continuous speech recognition is to assume a simple probabilistic model of speech production whereby a specified word sequence,  $W$ , produces an acoustic observation sequence  $Y$ , with probability  $P(W, Y)$ . The goal is then to decode the word string, based on the acoustic observation sequence, so that the decoded string has the maximum a posteriori (MAP) probability.” Lawrence Rabiner & Biing-Hwang Huang, FUNDAMENTALS OF SPEECH RECOGNITION at 434 (1993). “The way in which we compute  $P(Y|W)$ , for large vocabulary speech recognition, is to build statistical models for subword speech units, build up word models from these subword speech unit models (using a lexicon to describe the composition of words), and then postulate word sequences and evaluate the acoustic model probabilities via standard concatenation methods.” *Id.* at 435.

One of the most common probabilistic techniques for speech recognition was Hidden Markov Models (HMMs), also sometimes referred to as “Markov sources” or “probabilistic functions of Markov chains.” *See id.* at 321–322. “The underlying assumption of the HMM . . . is that the speech signal can be well characterized as a parametric random process, and that the parameters of the stochastic process can be determined (estimated) in a precise, well-defined manner.” *Id.* at 322. In simple terms, HMMs are a probabilistic technique of modeling (i.e.,

estimating) the likelihood (i.e., probability) that a given sequence of acoustic observations was produced by the utterance of a given speech unit (e.g., a word). For example, “consider the following simple isolated-word speech recognizer. For each word of a  $W$  word vocabulary, we want to design a separate  $N$ -state HMM. We represent the speech signal of a given word as a time sequence of coded spectral vectors. We assume that the coding is done using a spectral codebook with  $M$  unique spectral vectors; hence each observation is the index of the spectral vector closest (in some spectral distortion sense) to the original speech signal. Thus, for each vocabulary word, we have a training sequence consisting of a number of repetitions of sequences of codebook indices of the word (by one or more talkers). The first task is to build individual word models. This task is done by using the solution to Problem 3 to optimally estimate model parameters for each word model. . . . [O]nce the set of  $W$  HMMs has been designed and optimized, recognition of an unknown word is performed using the solution to Problem 1 to score each word model based upon the given test observation sequence, and select the word whose model score is highest (i.e., the highest likelihood).” *Id.* at 333–334. In other words, in this simple example of an isolated-word speech recognizer, each word in the vocabulary is represented by a HMM that encodes the likelihood (i.e., probability) that a given sequence of acoustic observations was produced by the utterance of that word. For example, assuming the parameters of the HMMs have been appropriately estimated, the HMM for the word “vox” should produce a high probability (e.g., score) in response to the spoken utterance 'våks, with the HMM for “box” coming in close second.

By the late 1990s it was well-known and widely understood by POSITAs that HMMs “provide[] a natural and highly reliable way of recognizing speech for a wide range of applications and integrate[] well into systems incorporating both task syntax and semantics.” *Id.* at 322; *see also* H. Singer, *An Overview of Speech Recognition with Applications for Medical Professionals*,

BME Vol. 12, No. 11, pp. 27–32 (1999), available at [https://www.jstage.jst.go.jp/article/jsmbe1987/13/11/13\\_11\\_27/article](https://www.jstage.jst.go.jp/article/jsmbe1987/13/11/13_11_27/article) (“The most successful and widely used acoustic models in recent years have been Hidden Markov Models (HMM). Practically all major speech recognition systems are implemented using some sort of HMM.”). Indeed, by that time, the effective use of HMMs for speech recognition had been known and well-understood for several decades—“[t]he basic theory of hidden Markov models was published in a series of classic papers by Baum and his colleagues ([1]-[5]) in the late 1960s and early 1970s and was implemented for speech-processing applications by Baker [6] at CMU, and by Jelinek and his colleagues at IBM ([7]-[13]) in the 1970s.” Lawrence Rabiner & Biing-Hwang Huang, FUNDAMENTALS OF SPEECH RECOGNITION at 322 (1993) (citing, *inter alia*, James K. Baker, *The DRAGON System—An Overview*, IEEE Trans. Acoustics, Speech, Signal Proc., Vol. ASSP-23, No. 1, pp. 24–29 (Feb. 1975)). The FUNDAMENTALS OF SPEECH RECOGNITION text, for example, devotes an entire chapter to the history, principles, design, and application of HMMs to speech recognition. *See id.*, ch. 6.

### **C. Using Multimodal Input with Speech Recognition Technology**

By no later than the late 1990s, it was further well understood that multimodal inputs could be used with speech recognition technology. For example, the Multimodal Maps System (“MMAPS”), which was publicly available in the United States at least as early as 1997, allowed the user to use “drawn gestures and written natural language statements” in combination with natural language spoken input. Cheyer and Julia, *Multimodal Maps: An Agent-Based Approach* (1995) at 1. More specifically, requests made by the user to MMAP could be “entered using gestures (e.g. arrows, lines, circles, cross-out or delete marks, etc.), handwriting, voice, or a combination of pen and voice.” *Id.* at 2. The system even permitted “the user to *simultaneously combine* direct manipulation, gestural drawings, handwritten, typed and spoken natural language.”

A. Cheyer & L. Julia, *Multimodal Maps: An Agent-Based Approach*, Proc. 1st Int’l Conf. Cooperative Multimodal Comm. (May 1995) (CMC/95), in: *Multimodal Human-Computer Communication*, Lecture Notes in Computer Science, Vol. 1374, at pp. 111-121 at 113. The various inputs were then combined “across all modalities to form a valid command for the application.” *Id.* at 117.

As another example, the AT&T MATCH System (“MATCH”), which was available publicly in the United States by December 2001 and no later than July 2004, allowed users to issue requests through a graphical interface using a combination of multimodal inputs. MATCH provided “a working city guide and navigation system” that enabled “mobile users to access restaurant and subway information for New York City.” M. Johnston, et al., *MATCH: An Architecture for Multimodal Dialogue Systems* (2002) at 1. Users could “give commands using speech, by drawing on the display with a stylus or using synchronous multimodal combinations of the two modes.” M. Johnston, et al., *MATCH: Multimodal Access To City Help* (2001) at 1.



Fig. 2. MATCH user interface

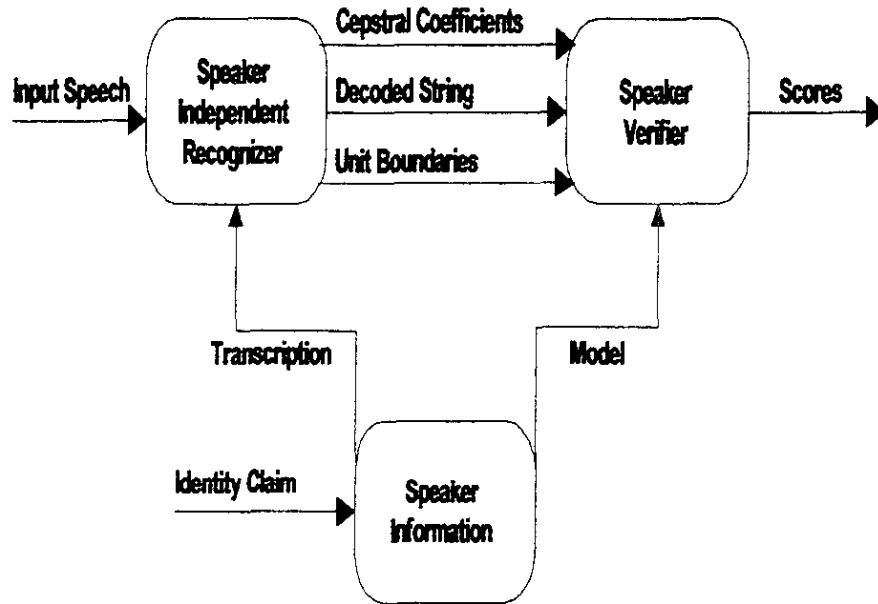
*Id.* at Fig. 2. Using the MATCH user interface, an example of which is shown above, “a user might ask to see cheap Italian restaurants in Chelsea by saying *show cheap Italian restaurants in Chelsea*, by circling an area on the map and saying *show cheap Italian restaurants in this neighborhood*; or, in a noisy or public environment, by circling an area and writing *cheap and Italian*.” M. Johnston, et al., *MATCH: An Architecture for Multimodal Dialogue Systems* (2002) at 1–2.

#### **D. Accessing Information with Domain Agents**

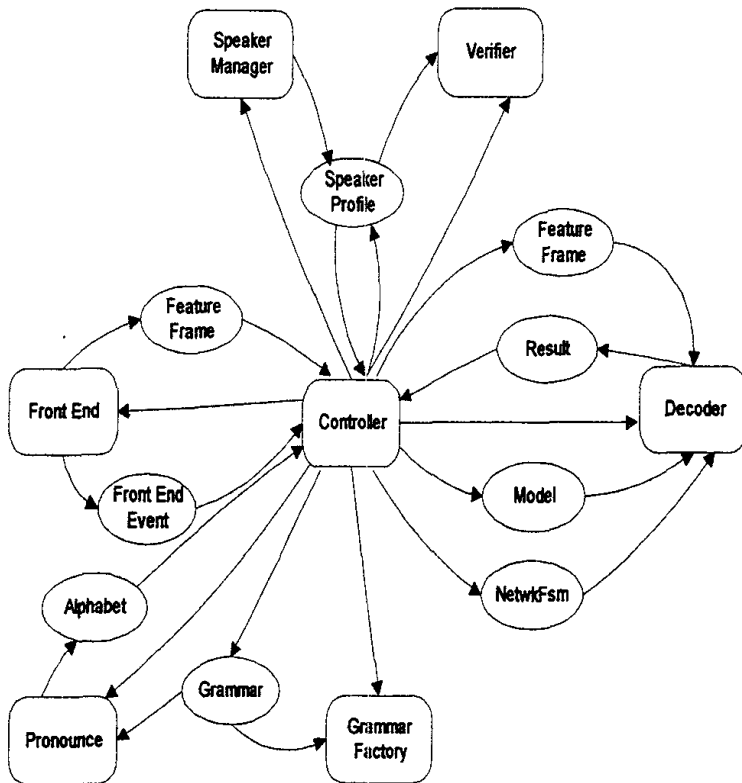
The idea of utilizing domain agents to access information was well known in the art by the mid-1990s and would have been within the knowledge of one of skill in art prior to the priority date of the Asserted patents. One example of domain agents can be found in the MMAPS system, which utilized the concept of a domain agent to facilitate “distributed access to commercial applications, such as mail systems, calendar programs, databases, etc.” Cheyer and Julia, *Multimodal Maps: An Agent-based Approach* (1995) at 2. Specifically, MMAPS employed an “agent-based approach” through its implementation of its “Open Agent Architecture,” where the “agents are distributed entities that can run on different machines, and communicate together to solve a task for the user.” *Id.* at 3. Indeed, the “architecture for the OAA” system was based on the 1993 system, Schwartz’s FLiPSiDE, as the MMAPS creators have explained. *Id.* One kind of agent that the MMAPS system included was a “Domain Agent,” which was described as “encod[ing] knowledge as to what actions must be performed” to resolve requests for particular domains. *Id.* at 4. The Domain Agent is further described as follows: “Once arguments to a query [were] resolved, the domain agent coordinates the actions and calculations necessary to produce the result of the request.” *Id.* U.S. Patents 6,859,800 and 6,415,256 each also disclose domain agents, the former referring to this concept as “modules,” *see, e.g.*, 800 patent at 28:36-29:38, and the latter referring to this concept as “application programs,” *see, e.g.*, Fig. 3, element [20].

### **E. Using User Personalization Models with Speech Recognition Technology**

By no later than the late-1990s, it was also well understood to perform speech recognition using user personalization models based upon a user or plurality of users' previous interactions with the system. For example, AT&T's Watson Speech Recognition Engine, which was publicly available at least as of 1997, was "a software product that supports speech recognition, text to speech synthesis, and speaker verification." R. D. Sharp *et al.*, "The Watson speech recognition engine," 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany (1997) at 4065. The Watson engine made use of a speaker verification process through which "[a] user makes an identity claim by speaking a digit string or a phrase unique to that user." *Id.* at 4067. "[T]he recognized utterance functions as an identity claim and is used to retrieve a set of speaker-dependent models (created at enrollment), which is then transmitted to the verifier along with the segmented utterance." *Id.* The speaker-dependent models are based on a user's previous interactions with the system: "The system requires limited enrollment data (usually a few repetitions of the password phrase) in a single session and yet provides robust performance under a variety of test conditions that are mismatched with training. This is achieved by using likelihood ratio scoring and also by performing ongoing adaptation of speaker dependent models and thresholds using all accepted in-service verification utterances." *Id.* A block diagram of the speaker verification module, including a user model and repository for speaker information is shown below:



*Id.* The Watson Engine integrated the user model and speaker verification module into a broader architecture to perform speech recognition:



*Id.* at 4065; *see also id.* (“The SpeakerManager subsystem maintains a set of SpeakerProfiles in memory and or on disk, manages SpeakerProfiles for all known speakers and keeps track of the current speaker name and group. The Verifier subsystem is responsible for verifying that a set of utterances comes from a particular speaker.”).

As another example, the Dragon NaturallySpeaking System, which was publicly available and for sale by August 2000, enabled users to create profiles and provide input data that informed a personalized user model: “Each person who wants to use the program needs to create a new set of user speech files and train Dragon NaturallySpeaking to understand his or her voice. A single person might want to have more than one set of user speech files (called a user) if, for example, you dictate into a mobile recorder or in more than one language.” *See* Lernout & Hauspie and Dragon Systems, *Dragon NaturallySpeaking 5: User’s Guide* (August 2000) at 131; *see also id.* (“What are user speech files? Your user speech files contain all the information that Dragon NaturallySpeaking gathers about you: your pronunciation, your vocabulary, how often you use certain words, and your preferences (whether you want one or two spaces after a period or a full stop, for example).”). The Dragon Naturally Speaking System made use of user models and personalization to inform its speech recognition process: “Dragon NaturallySpeaking uses acoustic data about the sound patterns that make up different words to choose the words that most closely match what it heard. Since no two people sound exactly alike, Dragon NaturallySpeaking does a much better job of matching sounds to words when it knows something about your pronunciation. When you first trained the program, you provided acoustic data on top of what Dragon NaturallySpeaking already knows about the sounds of English.” *Id.* at 10.

Beyond these examples, numerous other publications, systems and references made use of

user personalization models to inform the automatic speech recognition process. *See, e.g.*, U.S. Patent App. Pub No. 2002/0087315 (Lee); S. Furui, K. Iwano, C. Hori, T. Shinozaki, Y. Saito and S. Tamura, “Ubiquitous speech processing,” 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings, 2001; Reithinger, Norbert, Jan Alexandersson, Tilman Becker, Anselm Blocher, Ralf Engel, Markus Löckelt, Jochen Müller et al. “Smartkom: adaptive and flexible multimodal access to multiple applications.” In Proceedings of the 5th international conference on Multimodal interfaces, 2003.

### **III. IDENTIFICATION OF PRIOR ART**

Defendant identifies below each item of prior art that invalidates each Asserted Claim of the Asserted Patents. All of the prior art cited and discussed below and in the attached charts anticipates and/or renders obvious, alone or in combination with the below-disclosed references, one or more of the Asserted Claims of the Asserted Patents. Exemplary combinations of prior art references that render obvious the Asserted Claims of the Asserted Patents are set forth in the attached Appendix A. Defendant expressly reserves the right to supplement and/or amend these Preliminary Invalidity Contentions as noted above.

Defendant also attaches limitation-by-limitation claim charts demonstrating where each limitation of the Asserted Claims of the Asserted Patents is found in the prior art. These claim charts are based on Defendant’s current understanding of the terms and limitations of the Asserted Claims, and its current understanding of the claim constructions apparently underlying Dialect’s Infringement Contentions (to the extent any such understanding is possible given Dialect’s inscrutable and boilerplate contentions). To the extent any limitation is deemed not to be met, explicitly or inherently, by an item of prior art, Defendant contends that the difference would have been obvious to a person of ordinary skill in the art and within the knowledge of one skilled in the art at the time of the alleged invention, so that the claimed invention would have been obvious

both in light of the single reference alone and/or in light of combined references.

Furthermore, Defendant identifies each of the prior art items disclosed and/or cited in these Preliminary Invalidity Contentions, the attached Exhibits, including the limitation-by-limitation claim charts attached hereto, for each Asserted Patent.

**A. Background Prior Art**

<b>Issued Patents and Published Patent Applications</b>
U.S. Patent No. 7,890,324 (Bangalore324)
U.S. Patent App. Pub No. 2003/0093419 (Bangalore419)
U.S. Patent No. 7,177,815 (Ehlen)
U.S. Patent App. Pub. No. 2002/0178005 A1 (Dusan)
U.S. Patent. No. 5,177,685 (Davis)
U.S. Patent No. 5,649,103 (Datta)
U.S. Patent No. 5,897,616 (Kanevsky)
U.S. Patent No. 5,946,654 (Newman)
U.S. Patent No. 6,029,124 (Gillick)
U.S. Patent No. 6,078,914 (Redfern)
U.S. Patent No. 6,157,912 (Kneser)
U.S. Patent No. 6,321,196 (Franceschi)
U.S. Patent No. 6,430,531 (Polish)
U.S. Patent No. 6,499,013 (Weber)
U.S. Patent No. 6,523,061 (Halverson)
U.S. Patent No. 6,052,657 (Yamron)
U.S. Patent No. 6,839,669 (Gould)
U.S. Patent No. 4,829,576 (Porter)
U.S. Patent No. 5,799,279 (Gould279)
U.S. Patent No. 5,027,406 (Roberts)
U.S. Patent No. 6,212,498 (Sherwood)

U.S. Patent No. 7,085,716 (Van Even)
U.S. Patent No. 7,092,928 (Elad)
U.S. Patent No. 7,216,073 (Lavi)
U.S. Patent No. 7,219,073 (Taylor)
U.S. Patent No. 8,392,188 (Riccardi)
WO1999046763 (Wilson)
WO2000021075 (Maes)
WO2001080096 (Khan)
<b>Other Publications</b>
SAM: A Perceptive Spoken Language Understanding Robot, Michael K. Brown, Bruce M. Buntschuh, and Jay G. Wilpon, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS 1992
Levin, E. and Pieraccini, R., "CHRONUS: the next generation", Proc. of 1995 ARPA Spoken Language Systems Technical Workshop, Austin, Texas, Jan. 1995
Automated Call Routing in a Telecommunications Network, A.L. Gorin, H. Haiak, R. Rose and L. Miller (1994)
On automated language acquisition, Allen Gorin (1995)
"How may I help you?" A.L. Gorin, G. Riccardi, J.H. Wright (1997)
S. Furui, K. Iwano, C. Hori, T. Shinozaki, Y. Saito and S. Tamura, "Ubiquitous speech processing," 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings, Salt Lake City, UT, USA, 2001
Reithinger, Norbert, Jan Alexandersson, Tilman Becker, Anselm Blocher, Ralf Engel, Markus Löckelt, Jochen Müller et al. "Smartkom: adaptive and flexible multimodal access to multiple applications." In Proceedings of the 5th international conference on Multimodal interfaces. 2003.
Bangalore, Srinivas, and Michael Johnston. "Integrating multimodal language processing with speech recognition." <i>Proc. ICSLP 2000</i> . 2000.
Turunen, Markku, et al. "Robust and adaptive architecture for multilingual spoken dialogue systems." <i>INTERSPEECH</i> . 2004.
M.D. Sadek, A. Ferrieux, A. Cozannet, P. Bretier, F. Panaget, and J. Simonin. EFFECTIVE HUMAN-COMPUTERCOOPERATIVE SPOKEN DIALOGUE: THE AGS DEMONSTRATOR (1996)
Candace A. Kamm Diane J. Litman Marilyn A. Walker. FROM NOVICE TO EXPERT: THE EFFECT OF TUTORIALS ON USER EXPERTISE WITH SPOKEN DIALOGUE SYSTEMS (1998)

Candace Kamm, et. al. "Evaluating Spoken Dialogue Systems for Telecommunication Services." (1997)
B. Buntschuh, C. Kamm, G. Di Fabbrizio, A. Abella, M. Mohri, S. Narayanan, I. Zeljkovic, R. D. Sharp, J. Wright, S. Marcus, J. Shaffer, R. Duncan and J. G. Wilpon. VPQ: A SPOKEN LANGUAGE INTERFACE TO LARGE SCALE DIRECTORY INFORMATION (1998)
R. Pieraccini, Esther Levin. "Stochastic Representation of Semantic Structure for Speech Understanding." (1991)
Alfred Kobsa, Jfirgen Allgayer, Carola Reddig, Norbert Reithinger Dagmar Schmauks, Karin Harbusch, Wolfgang Wahlster. Combining Deictic Gestures and Natural Language for Referent Identification. (1986)
B. M. Arons. "The audio-graphical interface to a personal integrated telecommunications system." (1984)
E. Levin, S. Narayanan, R. Pieraccini, K. Biatov, E. Bocchieri, G. Di Fabbrizio, W. Eckert, S. Lee, A. Pokrovsky, M. Rahim, P. Ruscitti, M. Walker. THE AT&T-DARPA COMMUNICATOR MIXED-INITIATIVE SPOKEN DIALOG SYSTEM (2000)
Esther Levin, Roberto Pieraccini, Wieland Eckert, Pino Di Fabbrizio, Shrikanth Narayanan. SPOKEN LANGUAGE DIALOGUE: FROM THEORY TO PRACTICE (1999)
Roberto Pieraccini, Esther Levin, Wieland Eckert. AMICA: the AT&T Mixed Initiative Conversational Architecture (1997)
Candace Kamm. User interfaces for voice applications (1995)
Glenn Rifkin. "Interface; A Phone that Plays Secretary for Travelers." (1994)
Chris Schmandt, et. al. "A Conversational Telephone Messaging System." (1984)
Richard A. Bolt. "Put-That-There": Voice and Gesture at the Graphics Interface (1980)
Christopher Schmandt, Eric A. Hulteen. The Intelligent Voice-Interactive Interface (1981)
Michael K. Brown, et. al. "A Context-Free Grammar Compiler for Speech Understanding Systems." (1994)
M.A. Walker, et. al. "Learning Optimal Dialogue Strategies: A Case Study of a Spoken Dialogue Agent for Email." (1998)
Joseph Polifroni and Stephanie Seneff. "Galaxy-II as an Architecture for Spoken Dialogue Evaluation." (2000)
Giuseppe Di Fabbrizio, Dawn Dutton, Narendra Gupta, Barbara Hollister, Mazin Rahim, Giuseppe Riccardi, Robert Schapire and Juergen Schroeter. AT&T HELP DESK. (2002)
James R. Glass, Timothy J. Hazen, and I. Lee Hetherington. REAL-TIME TELEPHONE-BASED SPEECH RECOGNITION IN THE JUPITER DOMAIN (1999)
Victor Zue, et. al. JUPITER: A Telephone Based Conversational Interface for Weather Information (2000)
Xiaolong Mou, Stephanie Seneff, Victor Zue. "Context-dependent Probabilistic Hierarchical Sub-lexical Modelling Using Finite State Transducers." (2001)
Lauren Baptist and Stephanie Seneff. GENESIS-II: A VERSATILE SYSTEM FOR LANGUAGE GENERATION IN CONVERSATIONAL SYSTEM APPLICATIONS (2000)
Stephanie Seneff. TINA: A Natural Language System for Spoken Language Applications

(1992)
Stephanie Seneff, Ed Hurley, Raymond Lau, Christine Pao, Philipp Schmid, and Victor Zue. GALAXY-II: A REFERENCE ARCHITECTURE FOR CONVERSATIONAL SYSTEM DEVELOPMENT (1998)
Jody J. Daniels and Benjamin Bell. "Listen-Communicate-Show (LCS): Spoken Language Command of Agent-based Remote Information Access." (2000)
Jody Daniels. Integrating a Spoken Language System with Agents for Operational Information Access (2000)
L. Bahl and F. Jelinek, "Decoding for channels with insertions, deletions, and substitutions with applications to speech recognition," in IEEE Transactions on Information Theory, vol. 21, no. 4, pp. 404-411, July 1975, doi: 10.1109/TIT.1975.1055419.
Jelinek, Frederick, Lalit Bahl, and Robert Mercer. "Design of a linguistic statistical decoder for the recognition of continuous speech." IEEE Transactions on Information Theory 21, no. 3 (1975): 250-256.
F. Jelinek, "Continuous speech recognition by statistical methods," in Proceedings of the IEEE, vol. 64, no. 4, pp. 532-556, April 1976, doi: 10.1109/PROC.1976.10159.
Bahl, Lalit R., Frederick Jelinek, and Robert L. Mercer. "A maximum likelihood approach to continuous speech recognition." IEEE transactions on pattern analysis and machine intelligence 2 (1983): 179-190.
Lin, Bor-Shen & Wang, Hsin-min & Fen, Qin. (1999). A Distributed Architecture For Cooperative Spoken Dialogue Agents With Coherent Dialogue State And History.
G. Chung, S. Seneff and I.L. Hetherington, "Towards Multi-Domain Speech Understanding Using a Two-Stage Recognizer," Proc. Eurospeech, Budapest, Hungary, September 1999.
Maes, S.H. (2000) Elements of conversational computing - a paradigm shift. Proc. 6th International Conference on Spoken Language Processing (ICSLP 2000), vol. 1, 130-133, doi: 10.21437/ICSLP.2000-32
Higashida, M., Ohmori, K. (2000) A new dialogue control method based on human listening process to construct an interface for ascertaining a user's inputs. Proc. 6th International Conference on Spoken Language Processing (ICSLP 2000), vol. 1, 508-511, doi: 10.21437/ICSLP.2000-126
Utilizing statistical dialogue act processing in VERBMOBIL Authors: Norbert Reithinger, Elisabeth Maier ACL '95: Proceedings of the 33rd annual meeting on Association for Computational Linguistics Pages 116 - 121 <a href="https://doi.org/10.3115/981658.981674">https://doi.org/10.3115/981658.981674</a>
Paul Bamberg and Laurence Gillick. 1990. Phoneme-in-context modeling for dragon's continuous speech recognizer. In Proceedings of the workshop on Speech and Natural Language (HLT '90). Association for Computational Linguistics, USA, 163-169. <a href="https://doi.org/10.3115/116580.116635">https://doi.org/10.3115/116580.116635</a>
Wegmann, S., Zhan, P., Carp, I., Newman, M., Yamron, J., Gillick, L. (1999) Dragon systems' 1998 broadcast news transcription system. Proc. 6th European Conference on Speech Communication and Technology, 691-694, doi: 10.21437/Eurospeech.1999-177x
Zadrozny, Wlodek, Marcin Szummer, Stanislaw Jarecki, David E Johnson, and Leora Morgenstern. 1994. "NL Understanding with a Grammar of Constructions." Proc. Coling'94. 1994.
S. Seneff, D. Goddeau, C. Pao and J. Polifroni, "Multimodal discourse modelling in a multi-user

multi-domain environment,” Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96, Philadelphia, PA, USA, 1996, pp. 192-195 vol.1, doi: 10.1109/ICSLP.1996.607074
Lin, Bor-shen, Hsin-min Wang, and Lin-shan Lee. “A distributed architecture for cooperative spoken dialogue agents with coherent dialogue state and history.” In Asru, vol. 99, p. 4. 1999.
Kawaguchi, Nobuo, Shigeki Matsubara, Katsuhiko Toyama, and Yasuyoshi Inagaki. “An architecture for multi-domain spoken dialog systems.” In Proceedings of the 5th Natural Language Processing Pacific Rim Symposium (NLPRS'99), pp. 463-466. 1999.
<b>Systems</b>
AT&T How May I Help You?
AT&T Help Desk
AT&T ANNIE
AT&T AMICA
AT&T Communicator
AT&T Voice Post Query
MIT Galaxy II
MIT Jupiter
MIT Genesis
MIT TINA
Wildfire Assistant
MIT Phone Slave
AGS Demonstrator
Lockheed Martin Listen-Communicate-Show System

**B. Charts Setting Forth Where in the Prior Art References Each Element of the Asserted Claim(s) Are Found**

Attached to these Preliminary Infringement Contentions are limitation-by-limitation claim charts demonstrating where specifically each limitation of the Asserted Claims is found in the identified references. These claim charts are based on Defendant’s current understanding of the terms and limitations of the Asserted Claims, and its current understanding of the claim constructions apparently underlying Dialect’s infringement contentions (to the extent any such understanding is possible given Dialect’s inscrutable and boilerplate contentions). To the extent

any limitation is deemed not to be met explicitly or inherently by an item of prior art, Defendant contends that the difference would have been obvious to a person of ordinary skill in the art and within the knowledge of one skilled in the art at the time of the alleged invention, so that the claimed invention would have been obvious both in light of the single reference alone and/or in light of combined references.

As a general matter, all portions of each disclosed item of prior art are relied upon to support the disclosure of each Asserted Claim limitation, as all portions provide general support. Supporting citations are nevertheless provided, but do not necessarily represent every location where a particular feature or claim element may be found in the prior art item. For system prior art, Defendant may rely on the personal knowledge, experience, and testimony of any relevant witnesses including their technical experts, as evidence of the operation of the prior art system and how the system meets each Asserted Claim limitation. Where elements are disclosed at multiple locations within a single item of prior art, Defendant has not necessarily identified each iteration of every disclosure. Defendant therefore reserves the right to rely on additional, or different, portions of the prior art items other than those specifically cited in the claim charts provided herewith. Defendant also reserves the right to supplement and/or amend these contentions based on the herein-identified prior art considered individually or in combination with additional information learned during discovery and/or in light of the Court's claim construction rulings. In addition, Defendant may become aware of additional references as discovery progresses, and Defendant reserves the right to use such references to demonstrate that the Asserted Claims of the Asserted Patents are invalid as anticipated or obvious.

**1. '209 Patent**

<b>Reference</b>	<b>Name</b>	<b>Claim Chart(s)</b>
Colorado University Communicator System	CU Communicator	209-1
Carnegie Mellon University Communicator System	CMU Communicator	209-2
U.S. Patent No. 7,085,723	Ross	209-3
U.S. Patent No. 5,937,383	Ittycheriah	209-4
PCT Publication No. WO0011571	Hartono	209-5
U.S. Patent No. 5,680,511	Baker	209-6
U.S. Patent No. 7,249,018	Ross018	209-7
U.S. Patent No. 6,636,848	Aridor	209-8
Dragon Naturally Speaking System	Dragon	209-9
SRI's ATIS System	SRI ATIS	209-10
IBM ViaVoice 9.0 System	ViaVoice	209-11
General Magic Portico System	Portico	209-12
Tellme Networks System	Tellme	209-13
U.S. Patent No. 6,829,603	Chai	209-14
Concept-Based Search and User Modeling in Information Retrieval Based on Human-Machine Dialogue	Abe	209-15

**2. '006 Patent**

<b>Reference</b>	<b>Name</b>	<b>Claim Chart(s)</b>
Colorado University Communicator System	CU Communicator	006-1
Carnegie Mellon University Communicator System	CMU Communicator	006-2
U.S. Patent No. 7,085,723	Ross	006-3
U.S. Patent No. 5,937,383	Ittycheriah	006-4

<b>Reference</b>	<b>Name</b>	<b>Claim Chart(s)</b>
PCT Publication No. WO0011571	Hartono	006-5
U.S. Patent No. 5,680,511	Baker	006-6
U.S. Patent No. 7,249,018	Ross018	006-7
U.S. Patent No. 6,636,848	Aridor	006-8
Dragon Naturally Speaking System	Dragon	006-9
SRI's ATIS System	SRI ATIS	006-10
IBM ViaVoice 9.0 System	ViaVoice	006-11
Spoken Language Processing: A Guide to Theory, Algorithm, and System Development	Acero	006-12
U.S. Patent No. 6,829,603	Chai	006-13
Concept-Based Search and User Modeling in Information Retrieval Based on Human-Machine Dialogue	Abe	006-14

### 3. '825 Patent

<b>Reference</b>	<b>Name</b>	<b>Claim Chart(s)</b>
Colorado University Communicator System	CU Communicator	825-1
Carnegie Mellon University Communicator System	CMU Communicator	825-2
U.S. Patent No. 7,085,723	Ross	825-3
U.S. Patent No. 5,937,383	Ittycheriah	825-4
PCT Publication No. WO0011571	Hartono	825-5
U.S. Patent No. 5,680,511	Baker	825-6
U.S. Patent Application Publication No. 2002/0165860	Glover	825-7
Dragon Naturally Speaking System	Dragon	825-8
SRI's ATIS System	SRI ATIS	825-9
IBM ViaVoice 9.0 System	ViaVoice	825-10

Reference	Name	Claim Chart(s)
U.S. Patent No. 7,283,964	Geilhufe	825-11
U.S. Patent No. 6,785,651	Wang	825-12
U.S. Patent No. 6,829,603	Chai	825-13
Concept-Based Search and User Modeling in Information Retrieval Based on Human-Machine Dialogue	Abe	825-14

#### 4. '607 Patent

Reference	Name	Claim Chart(s)
IBM ViaVoice 9.0 System	ViaVoice	607-1
Dragon Naturally Speaking System	Dragon	607-2
AT&T MATCH System	MATCH	607-3
U.S. Patent No. 6,964,023	Maes	607-4
U.S. Patent No. 6,415,256	Ditzik	607-5
U.S. Patent App. Pub No. 2002/0133354	Ross	607-6
U.S. Patent App. Pub No. 2002/0087315	Lee	607-7
MMAPs System	MMAPs	607-8
WO2000020962	Coffman	607-9
U.S. Patent App. Pub No. 2004/0044516	Kennewick	607-10
U.S. Patent No. 5,937,383	Ittycheriah	607-11
Effects of dialog initiative and multi-modal presentation strategies on large directory information access	Narayanan	607-12
U.S. Patent App. Pub No. 2004/0120472A1	Popay	607-13

## 5. '039 Patent

Reference	Name	Claim Chart(s)
U.S. Patent No. 5,748,974	Johnson	039-1
A Framework for Multi-Agent Systems with Multi-Modal User Interfaces in Distributed Computing Environments	Park	039-2
Can Many Agents Answer Questions Better Than One?	Galitsky	039-3
U.S. Patent No. 6,964,023	Maes	039-4
U.S. Patent No. 7,447,635	Konopka	039-5
U.S. Patent Pub. No. 2002/0133354	Ross	039-6
WO2000011571	Hartono	039-7
AT&T MATCH System	MATCH	039-8
MMAPS System	MMAPs	039-9
Dragon Naturally Speaking System	Dragon	039-10
An Architecture for Multi-Domain Spoken Dialog Systems	Kawaguchi	039-11
Effects of Dialog Initiative and Multi-Modal Presentation Strategies on Large Directory Information Access	Narayanan	039-12

### C. Obviousness

With regard to the motivation to combine cited references under 35 U.S.C. § 103, as the United States Supreme Court held in *KSR International Co. v. Teleflex, Inc.*, “[t]he combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results.” 550 U.S. 398, 416 (2007). The Supreme Court further held that, “[w]hen a work is available in one field of endeavor, design incentives and other market forces can prompt variations of it, either in the same field or a different one. If a person of ordinary skill can implement a predictable variation, § 103 likely bars its patentability. For the same reason, if

a technique has been used to improve one device, and a person of ordinary skill in the art would recognize that it would improve similar devices in the same way, using the technique is obvious unless its actual application is beyond his or her skill.” *Id.* at 417.

Moreover, the Supreme Court has held that “in many cases a person of ordinary skill will be able to fit the teachings of multiple patents together like pieces of a puzzle.” *Id.* at 420. Indeed, the Supreme Court held that it is sufficient that a combination of elements was “obvious to try” holding that, “[w]hen there is a design need or market pressure to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within his or her technical grasp. If this leads to the anticipated success, it is likely the product not of innovation but of ordinary skill and common sense. In that instance the fact that a combination was obvious to try might show that it was obvious under § 103.” *Id.* at 421. “Rigid preventative rules that deny factfinders recourse to common sense, however, are neither necessary under our case law nor consistent with it.” *Id.*

Finally, the Supreme Court recognized that “[g]ranted patent protection to advances that would occur in the ordinary course without real innovation retards progress and may, in the case of patents combining previously known elements, deprive prior inventions of their value or utility.” *Id.* at 419.

For the Asserted Patents, each of the prior art items disclosed in these Preliminary Invalidity Contentions is directed to the same or similar fields of endeavor. Accordingly, one of ordinary skill in the art would have recognized that the results of combining these items of prior art (e.g., disclosures therein) were predictable and would have been clearly motivated to combine any of the prior art items identified below and in the attached charts for each Asserted Patent, including to make any necessary modifications, to arrive at the alleged inventions of the Asserted

Claims. Further, Dialect has not identified, and could not identify, any unexpected results that ensued from combining familiar elements according to known techniques.

Accordingly, at least the following rationales support a finding of obviousness based on the combination of each prior art reference disclosed with one or more of any of the other prior art references disclosed:

- (A) Combining prior art elements according to known methods to yield predictable results;
- (B) Simple substitution of one known element for another to obtain predictable results;
- (C) Use of known technique to improve similar devices (methods, or products) in the same way;
- (D) Applying a known technique to a known device (method, or product) ready for improvement to yield predictable results;
- (E) “Obvious to try”—choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success;
- (F) Known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations would have been predictable to one of ordinary skill in the art; and
- (G) Some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

Thus, Defendant contends that the Asserted Claims are further invalid as obvious under 35 U.S.C. § 103 in view of at least the prior art identified herein. Each reference identified herein can and would be combined with any of the other identified references so as to render the Asserted Claims. Accordingly, if any reference is asserted or found not to disclose a particular element, Defendant reserves the right to identify and rely on combinations with other references that do disclose the purportedly missing element.

Similarly, to the extent that any identified reference is argued or found not to anticipate one or more Asserted Claims, then references described herein that disclose the purportedly

missing element, and/or the knowledge and skill of a person of ordinary skill in the art at the time of the invention that does so, can and would be combined with that reference to render the claim invalid as obvious.

### **1. '209 Patent**

Defendant incorporates by reference the preceding legal background and discussion of obviousness combinations. In addition, Defendant discloses additional motivations to combine regarding the asserted claims in the '209 patent. Cited references CU Communicator, CMU Communicator, SRI ATIS, General Magic Portico, and TellMe ("Group 209-A") are telephone-based dialog systems that incorporate speech recognition. Ross, Ross018, and IBM ViaVoice constitute and/or disclose IBM's speech recognition systems and technology ("Group 209-B"). Baker and Dragon Naturally Speaking constitute and/or disclose Dragon System's speech recognition systems and technology ("Group 209-C"). Ittycheriah describes a speech recognition engine (collectively, with Group 209-B and Group 209-C, "Group 209-D"). Hartono describes a speech recognition agent adapted to receive natural language input ("Group 209-E"). Aridor describes methods for searching the Internet for information using multiple search engines simultaneously. Chai discusses methods for natural language based queries in a network environment. Abe describes information retrieval over the Internet using concept based search and spoken language input (collectively, with Aridor and Chai, "Group 209-F").

A person of ordinary skill in the art would have been motivated to combine Group 209-A with 209-D. Off-the-shelf speech engines could be used to convert speech to text. Therefore, combining the known speech recognition technology with the telephone based dialog systems merely combined prior art elements according to known methods to yield predictable results. It was also obvious to try, because a person of ordinary skill in the art could choose to incorporate into the telephone based dialog systems a finite number of identified, predictable solutions, with a

reasonable expectation of success.

A person of ordinary skill in the art would also have been motivated to combine any of the references composing Group 209-A with any other of those references. Each of those references are telephone-based dialog systems that incorporate speech recognition. Therefore, a person of ordinary skill in the art could make a simple substitution of one known element in one reference for a known element in another to obtain predictable results.

For the same reason, a person of ordinary skill in the art would also have been motivated to combine any of the references composing Group 209-D with any other of those references. Each of those references describe speech recognition systems and technology. Therefore, a person of ordinary skill in the art could make a simple substitution of one known element in one reference for a known element in another to obtain predictable results. Furthermore, a person of ordinary skill in the art would have been motivated to combine any of the references composing Group 209-B and Group 209-C with any other of those references. Group 209-B constitutes and/or discloses IBM's speech recognition systems and technology. Group 209-C constitutes and/or discloses Dragon System's speech recognition systems and technology. Therefore, each of those groups contains a teaching, suggestion, or motivation that would have led one of ordinary skill to use compatible IBM and Dragon systems with other compatible IBM and Dragon systems, respectively.

A person of ordinary skill in the art would have been motivated to combine Group 209-A and Group 209-D with Group 209-E. Hartono discloses that “[a]ny suitable off the shelf speech recognition systems may be used as the automatic speech recognition system 102a.” 11:4-12. A person of ordinary skill reading Hartono would thus be motivated to look for available speech recognition systems in Group 209-A and Group 209-D to use to with Hartono.

A person of ordinary skill in the art would have been motivated to combine Group 209-A and Group 209-D with Group 209-F. By 2002, a person of ordinary skill in the art would have understood that the Internet provided a robust source of information for information retrieval in response to queries. Moreover, voice recognition technology was already integrated with internet and network search technology to furnish information in response to queries. Therefore, a person of ordinary skill would have known to use the known technique of internet or network search to improve voice recognition technology in the same way.

Additionally, a person of ordinary skill in the art would have been motivated to combine any of the references in Groups 209-A, 209-B, 209-C, 209-D, 209-E, and 209-F with any other reference in any of those groups. All of those references come from the same fields of speech recognition and/or information retrieval. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these

references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference arrive at the claimed invention.

A person of ordinary skill in the art would also have been motivated to combine any of the references in Groups 209-A, 209-B, 209-C, 209-D, 209-E, and 209-F with any other reference discussed in the Background Section 2(a) above concerning existing speech recognition references. All of those references come from the same fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition would have made use of those references obvious to try and provide a reasonable expectation of success. As inventor Michael Kennewick has previously testified at trial in *VB Assets v. Amazon* matter, speech recognition components were a “commodity product.” Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of

these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

A person of ordinary skill in the art would have been motivated to combine any of the references in Groups 209-A, 209-B, 209-C, 209-D, 209-E, and 209-F with any other reference discussed in the Background Section 2(b) above concerning existing prior probability references. All of those references describe use of prior probabilities in the same general field (speech recognition) and/or information retrieval, and the rich history of development in the field of speech recognition generally and the long standing use of prior probability models within that field would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with

another reference.

Finally, a person of ordinary skill in the art would have been motivated to combine any of the references in Groups 209-A, 209-B, 209-C, 209-D, 209-E, and 209-F with any other reference discussed in the Background Section 2(d) above concerning existing domain agent references. All of those references describe the use of domain agents in the same fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition and the long standing use of domain agents would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

## 2. '006 Patent

Defendant incorporates by reference the preceding legal background and discussion of obviousness combinations. In addition, Defendant discloses additional motivations to combine regarding the asserted claims in the '006 patent. Cited references CU Communicator, CMU Communicator, and SRI ATIS ("Group 006-A") are telephone-based dialog systems that incorporate speech recognition. Ross, Ross018, and IBM ViaVoice constitute and/or disclose IBM's speech recognition systems and technology ("Group 006-B"). Baker and Dragon Naturally Speaking constitute and/or disclose Dragon System's speech recognition systems and technology ("Group 006-C"). Ittycheriah and Acero describe speech recognition engines (collectively, with Group 006-B and Group 006-C, "Group 006-D"). Hartono describes a speech recognition agent adapted to receive natural language input ("Group 006-E"). Aridor describes methods for searching the Internet for information using multiple search engines simultaneously. Chai discusses methods for natural language based queries in a network environment. Abe describes information retrieval over the Internet using concept based search and spoken language input (collectively, with Aridor and Chai, "Group 006-F").

A person of ordinary skill in the art would have been motivated to combine Group 006-A with 006-D. Off-the-shelf speech engines could be used to convert speech to text. Therefore, combining the known speech recognition technology with the telephone based dialog systems merely combined prior art elements according to known methods to yield predictable results. It was also obvious to try, because a person of ordinary skill in the art could choose to incorporate into the telephone based dialog systems a finite number of identified, predictable solutions, with a reasonable expectation of success.

A person of ordinary skill in the art would also have been motivated to combine any of the references composing Group 006-A with any other of those references. Each of those references

are telephone-based dialog systems that incorporate speech recognition. Therefore, a person of ordinary skill in the art could make a simple substitution of one known element in one reference for a known element in another to obtain predictable results.

For the same reason, a person of ordinary skill in the art would also have been motivated to combine any of the references composing Group 006-D with any other of those references. Each of those references describe speech recognition systems and technology. Therefore, a person of ordinary skill in the art could make a simple substitution of one known element in one reference for a known element in another to obtain predictable results. Furthermore, a person of ordinary skill in the art would have been motivated to combine any of the references composing Group 006-B and Group 006-C with any other of those references. Group 006-B constitutes and/or discloses IBM's speech recognition systems and technology. Group 006-C constitutes and/or discloses Dragon System's speech recognition systems and technology. Therefore, each of those groups contains a teaching, suggestion, or motivation that would have led one of ordinary skill to use compatible IBM and Dragon systems with other compatible IBM and Dragon systems, respectively.

A person of ordinary skill in the art would have been motivated to combine Group 006-A and Group 006-D with Group 006-E. Hartono discloses that “[a]ny suitable off the shelf speech recognition systems may be used as the automatic speech recognition system 102a.” 11:4-12. A person of ordinary skill reading Hartono would thus be motivated to look for available speech recognition systems in Group 006-A and Group 006-D to use to with Hartono.

A person of ordinary skill in the art would have been motivated to combine Group 006-A and Group 006-D with Group 006-F. By 2002, a person of ordinary skill in the art would have understood that the Internet provided a robust source of information for information retrieval in

response to queries. Moreover, voice recognition technology was already integrated with internet and network search technology to furnish information in response to queries. Therefore, a person of ordinary skill would have known to use the known technique of internet or network search to improve voice recognition technology in the same way.

Additionally, a person of ordinary skill in the art would have been motivated to combine any of the references in Groups 006-A, 006-B, 006-C, 006-D, 006-E, and 006-F with any other reference in any of those groups. All of those references come from the same fields of speech recognition and/or information retrieval. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference arrive at the claimed invention.

A person of ordinary skill in the art would also have been motivated to combine any of the references in Groups 006-A, 006-B, 006-C, 006-D, 006-E, and 006-F with any other reference discussed in the Background Section 2(a) above concerning existing speech recognition references. All of those references come from the same fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition would have made use of those references obvious to try and provide a reasonable expectation of success. As inventor Michael Kennewick has previously testified at trial in *VB Assets v. Amazon* matter, speech recognition components were a “commodity product.” Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

A person of ordinary skill in the art would have been motivated to combine any of the

references in Groups 006-A, 006-B, 006-C, 006-D, 006-E, and 006-F with any other reference discussed in the Background Section 2(b) above concerning existing prior probability references. All of those references describe use of prior probabilities in the same general field (speech recognition) and/or information retrieval, and the rich history of development in the field of speech recognition generally and the long standing use of prior probability models within that field would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

Finally, a person of ordinary skill in the art would have been motivated to combine any of the references in Groups 006-A, 006-B, 006-C, 006-D, 006-E, and 006-F with any other reference

discussed in the Background Section 2(d) above concerning existing domain agent references. All of those references describe the use of domain agents in the same fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition and the long standing use of domain agents would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

### **3. '825 Patent**

Defendant incorporates by reference the preceding legal background and discussion of obviousness combinations. In addition, Defendant discloses additional motivations to combine regarding the asserted claims in the '825 patent. Charted references CU Communicator, CMU

Communicator, and SRI ATIS (“Group 825-A”) are telephone-based dialog systems that incorporate speech recognition. Ross and IBM ViaVoice constitute and/or disclose IBM’s speech recognition systems and technology (“Group 825-B”). Baker and Dragon Naturally Speaking constitute and/or disclose Dragon System’s speech recognition systems and technology (“Group 825-C”). Ittycheriah and Geilhulfe describe a speech recognition engine (collectively, with Group 825-B and Group 825-C, “Group 825-D”). Hartono describes a speech recognition agent adapted to receive natural language input (“Group 825-E”). Glover describes methods for searching the Internet based on estimations of relevance. Chai and Wang discuss methods for natural language based queries in a network environment. Abe describes information retrieval over the Internet using concept based search and spoken language input (collectively, with Glover, Chai, and Wang, “Group 825-F”).

A person of ordinary skill in the art would have been motivated to combine Group 825-A with 825-D. Off-the-shelf speech engines could be used to convert speech to text. Therefore, combining the known speech recognition technology with the telephone based dialog systems merely combined prior art elements according to known methods to yield predictable results. It was also obvious to try, because a person of ordinary skill in the art could choose to incorporate into the telephone based dialog systems a finite number of identified, predictable solutions, with a reasonable expectation of success.

A person of ordinary skill in the art would also have been motivated to combine any of the references composing Group 825-A with any other of those references. Each of those references are telephone-based dialog systems that incorporate speech recognition. Therefore, a person of ordinary skill in the art could make a simple substitution of one known element in one reference for a known element in another to obtain predictable results.

For the same reason, a person of ordinary skill in the art would also have been motivated to combine any of the references composing Group 825-D with any other of those references. Each of those references describe speech recognition systems and technology. Therefore, a person of ordinary skill in the art could make a simple substitution of one known element in one reference for a known element in another to obtain predictable results. Furthermore, a person of ordinary skill in the art would have been motivated to combine any of the references composing Group 825-B and Group 825-C with any other of those references. Group 825-B constitutes and/or discloses IBM's speech recognition systems and technology. Group 825-C constitutes and/or discloses Dragon System's speech recognition systems and technology. Therefore, each of those groups contains a teaching, suggestion, or motivation that would have led one of ordinary skill to use compatible IBM and Dragon systems with other compatible IBM and Dragon systems, respectively.

A person of ordinary skill in the art would have been motivated to combine Group 825-A and Group 825-D with Group 825-E. Hartono discloses that “[a]ny suitable off the shelf speech recognition systems may be used as the automatic speech recognition system 102a.” 11:4-12. A person of ordinary skill reading Hartono would thus be motivated to look for available speech recognition systems in Group 825-A and Group 825-D to use to with Hartono.

A person of ordinary skill in the art would have been motivated to combine Group 825-A and Group 825-D with Group 825-F. By 2002, a person of ordinary skill in the art would have understood that the Internet provided a robust source of information for information retrieval in response to queries. Moreover, voice recognition technology was already integrated with internet and network search technology to furnish information in response to queries. Therefore, a person of ordinary skill would have known to use the known technique of internet or network search to

improve voice recognition technology in the same way.

Additionally, a person of ordinary skill in the art would have been motivated to combine any of the references in Groups 825-A, 825-B, 825-C, 825-D, 825-E, and 825-F with any other reference in any of those groups. All of those references come from the same fields of speech recognition and/or information retrieval. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference arrive at the claimed invention.

A person of ordinary skill in the art would also have been motivated to combine any of the references in Groups 825-A, 825-B, 825-C, 825-D, 825-E, and 825-F with any other reference discussed in the Background Section 2(a) above concerning existing speech recognition

references. All of those references come from the same fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition would have made use of those references obvious to try and provide a reasonable expectation of success. As inventor Michael Kennewick has previously testified at trial in *VB Assets v. Amazon* matter, speech recognition components were a “commodity product.” Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

A person of ordinary skill in the art would have been motivated to combine any of the references in Groups 825-A, 825-B, 825-C, 825-D, 825-E, and 825-F with any other reference discussed in the Background Section 2(b) above concerning existing prior probability references. All of those references describe use of prior probabilities in the same general field (speech

recognition) and/or information retrieval, and the rich history of development in the field of speech recognition generally and the long standing use of prior probability models within that field would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

Finally, a person of ordinary skill in the art would have been motivated to combine any of the references in Groups 825-A, 825-B, 825-C, 825-D, 825-E, and 825-F with any other reference discussed in the Background Section 2(d) above concerning existing domain agent references. All of those references describe the use of domain agents in the same fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition

and the long standing use of domain agents would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

#### **4. '607 Patent**

Defendant incorporates by reference the preceding legal background and discussion of obviousness combinations. In addition, Defendant discloses additional motivations to combine regarding the asserted claims in the '607 patent.

Charted references ViaVoice, Dragon, MATCH, MMAPs, Maes, and Ditzik (“Group 607-A”) constitute and/or disclose systems and technologies that, among other things, process speech and/or other inputs (including but not limited to text) from a user. Charted references Coffman, Ittycheriah, Ross and Narayanan (“Group 607-B”) describe systems and methods directed to,

among other things, determining the meaning of natural language user inputs and using contextually-relevant domain agents to return responsive results to users. Charted references Kennewick, Lee and Popay (“Group 607-C”) describe systems and methods that, among other things, utilize user information to process natural language inputs.

A person of ordinary skill in the art at the time of the alleged inventions of the Asserted Claims would have been motivated to combine the teachings of each of the references within and across each of Group 607-A, 607-B and 607-C because all of the references are directed to solving the same overarching problem of creating accurate, responsive and flexible conversational computing systems that generate contextually relevant responses to the user and are all in the same field of art. All of those references come from the same fields of natural language recognition and/or information retrieval. Known work in the field of natural language recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these

references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

It would have been obvious to combine any of the references in Group 607-A with any other reference in Group 607-A because all of those references are directed to processing multimodal natural language inputs. They apply known techniques—such as recognizing different forms of non-speech input, integrating non-speech input with speech inputs, and determining context from multiple forms of inputs—and are directed to solving the same problems for users. A skilled artisan would have understood that there are a finite number of predictable solutions for determining the meaning of user input and responding and/or taking action(s) based on the input, and thus, could have implemented the solutions disclosed by these systems in any of the other systems with a reasonable expectation of success. It would also have been obvious for a person of ordinary skill in the art to use compatible IBM (e.g., ViaVoice), AT&T (e.g., MATCH), and Dragon systems, with other compatible IBM, AT&T, and Dragon systems. Accordingly, skilled artisans would have been incentivized by market forces to incorporate the teachings of each of the other references into these systems.

It would have been obvious to combine any of the references in Group 607-B with any other reference in Group 607-B because all of those references are directed to using contextually-relevant domain agents to return responsive results to users. They apply known techniques—such as interpreting context from a user’s inputs and utilizing domain agents to generate responsive results—and are directed to solving the same problems for users. Combining these references would have required merely a simple substitution of known elements for processing these inputs. It would have been obvious to combine any of the references in Group 607-C with any other reference in Group 607-C because all of those references are directed to processing natural language user inputs

by making use of user-specific information. They apply known techniques—such as using user models and personalization profiles—and are directed to solving the same problems for users. Combining the teachings of these references would have been a matter of a skilled artisan applying known methods of natural language processing using user information and would have yielded predictable results.

It would have been obvious to combine any of the references in Group 607-A with any of the references in Group 607-B, and vice versa. Combining known multimodal natural language input processing systems and methods with known methods of using context and domain agents to generate responsive results to users merely combined prior art elements according to known methods to yield predictable results. It was also obvious to try, as a person of ordinary skill in the art could choose to incorporate into the multimodal natural language processing systems a finite number of identified, predictable solutions, with a reasonable expectation of success. As one example, MMAPS discloses the use of an “agent-based approach” in an “Open Agent Architecture,” where the “agents are distributed entities that can run on different machines, and communicate to solve a task for the user.” Cheyer and Julia, *Multimodal Maps: An Agent-based Approach* (1995) at 3; thus, a person of ordinary skill in the art would be motivated to use known methods for using contextually-relevant domain agents with a multimodal natural language processing system such as MMAPS. A person of ordinary skill in the art could make a simple substitution of one known element in one reference (e.g., recognition of one form of natural language inputs—speech) for a known element in another (e.g., multimodal natural language input processing) to obtain predictable results.

Additionally, it would have been obvious to combine Maes (Group 607-A) with Coffman and Ittycheriah (Group 607-B), as Maes expressly incorporates by reference Coffman and

Ittycheriah, all three references share a common named inventor, and all three have IBM as a common assignee. Furthermore, Ittycheriah discloses an example of a voice-controlled navigation application and Maes discloses an example implementation of an in-vehicle system. A person of ordinary skill in the art would be further motivated to combine the teachings of Maes and Ittycheriah because Ittycheriah's voice-controlled navigation would significantly benefit Maes's in-vehicle system by providing voice-based navigation assistance to drivers. In addition, a person of ordinary skill in the art would have been motivated to combine Coffman's teachings of organizing its context stack and updating it such that the most-recent context is at the top of the context with teachings of Maes and Ittycheriah because to do so would benefit Maes's system by providing efficient identification and retrieval of the most-recent context, allowing for faster executions and improving the responsiveness of Maes's system, and would additionally be inline with Ittycheriah's goals of quickly accepting or rejecting keywords.

It would have been obvious to combine any of the references in Group 607-A with any of the references in Group 607-C, and vice versa. Combining known multimodal natural language input processing systems and methods with known methods of using user-specific information to process natural language inputs (such as user models and user personalization profiles) merely combined prior art elements according to known methods to yield predictable results. It was also obvious to try, as a person of ordinary skill in the art could choose to incorporate into the multimodal natural language processing systems a finite number of identified, predictable solutions, with a reasonable expectation of success. As one example, MATCH utilized AT&T's Watson Speech Recognition Engine, which made use of "speaker verification" and "speaker dependent models" dependent on speaker-specific information. *See, e.g.*, R. D. Sharp et al., "The Watson speech recognition engine," 1997 IEEE International Conference on Acoustics, Speech,

and Signal Processing, Munich, Germany (1997) at 4067. As such, a person of ordinary skill in the art would be motivated to use known methods for utilizing user-specific information for processing natural language inputs in combination with a multimodal natural language processing system such as MATCH.

Finally, a person of ordinary skill in the art would have been motivated to combine any of the references in Groups 607-A, 607-B, and 607-C with any other reference in any of those groups. All of those references come from the same fields of natural language processing and/or information retrieval. Known work in the field of natural language processing would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference to arrive at the claimed invention.

Additionally, it would have been obvious to combine any of the references in Groups 607-

A, 607-B and 607-C with the background references noted above, *see supra* Sections III(A)-(B), for at least the reasons stated above and because the background references all disclose known techniques and methods such as recognizing words and phrases in spoken natural language utterances, using multimodal inputs with speech recognition technologies, accessing information with domain and using user personalization models. It would have been a matter of a simple substitution of these known techniques into the other prior art.

A person of ordinary skill in the art would have been motivated to combine any of the references in Groups 607-A, 607-B and 607-C with any other reference discussed in the Background Section 2(a) above concerning existing speech recognition references. All of those references come from the same fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition would have made use of those references obvious to try and provide a reasonable expectation of success. As inventor Michael Kennewick has previously testified at trial in *VB Assets v. Amazon* matter, speech recognition components were a “commodity product.” Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from

another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

A person of ordinary skill in the art would have been motivated to combine any of the references in Groups 607-A, 607-B and 607-C with any other reference discussed in the Background Section 2(c) above concerning existing multimodal references. All of those references describe the use of multiple input modes within the same general fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition and the long standing use of multiple modes of input into systems utilizing speech recognition would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from

the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

A person of ordinary skill in the art would have been motivated to combine any of the references in Groups 607-A, 607-B and 607-C with any other reference discussed in the Background Section 2(d) above concerning existing domain agent references. All of those references describe the use of domain agents in the same fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition and the long standing use of domain agents would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these

references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

A person of ordinary skill in the art would have been motivated to combine any of the references in Groups 607-A, 607-B and 607-C with any other reference discussed in the Background Section 2(e) above concerning existing user personalization references. All of those references describe using user personalization from the same fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition and the long standing use of user personalization with speech recognition models would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another

reference.

## 5. '039 Patent

Defendant incorporates by reference the preceding legal background and discussion of obviousness combinations. In addition, Defendant discloses additional motivations to combine regarding the asserted claims in the '039 patent.

Charted references Dragon, MATCH, MMAPs, Maes, Johnson, and Park (“Group 039-A”) constitute and/or disclose systems and technologies that, among other things, process speech and/or other inputs from a user. Charted references Galitsky, Ross, Hartono, and Narayanan (“Group 039-B”) describe systems and methods directed to, among other things, determining the meaning of natural language user’s inputs and using domain agents to return responsive results to users. Charted references Konopka and Kawaguchi (“Group 039-C”) describe speech recognition agents adapted to receive natural language input.

A person of ordinary skill in the art at the time of the alleged inventions of the Asserted Claims would have been motivated to combine the teachings of each of the references within and across each of Group 039-A, 039-B and 039-C because all of the references are directed to solving the same overarching problem of creating accurate, responsive and flexible conversational computing systems that generate contextually relevant responses to the user and are all in the same field of art.

It would have been obvious to combine any of the references in Group 039-A with any other reference in Group 039-A because all of those references are directed to processing multimodal natural language inputs and are thus in the same field of endeavor. They apply known techniques—such as recognizing speech inputs as well as different forms of non-speech input, integrating non-speech input with speech inputs, and determining context from multiple forms of inputs—and are directed to solving the same problems for users. A skilled artisan would have

understood that there are a finite number of predictable solutions for determining the meaning of user input and responding and/or taking action(s) based on the input, and thus, could have implemented the solutions disclosed by these systems in any of the other systems with a reasonable expectation of success. It also would have been obvious for a skilled artisan to combine compatible AT&T (e.g., MATCH) and Dragon systems with other compatible AT&T and Dragon systems. Additionally, to the extent these systems were commercialized, one of skill in the art would have been motivated to combine the existing technologies in the art so as to maximize the potential commercial success of the system by offering users a wide range of features, options, and use cases. Accordingly, skilled artisans would have been incentivized by market forces to incorporate the teachings of each of the other references into these systems.

For similar reasons, it would have been obvious to combine any of the references in Group 039-B with any other reference in Group 039-B because all of those references are in the same field of endeavor and employ the same kind of known techniques—such as interpreting context from a user’s inputs and utilizing domain agents to generate responsive results. Combining these references would have required merely a simple substitution of known elements for processing these inputs.

Likewise, it would have been obvious to combine any of the references in Group 039-C with any other reference in Group 039-C because those references are both in the field of speech recognition and systems. Combining the teachings of these references would have been a matter of a skilled artisan applying known methods of speech recognition and processing and would have yielded predictable results.

It would have been obvious to combine any of the references in Group 039-A with any of the references in Group 039-B, and vice versa. Combining the teachings of these references would

have simply resulted in applying known methods—like the backend processing elements of Group 039-B, including, but not limited to, domain agents—and would have resulted in predictable outcomes. Combining known multimodal natural language input processing systems and methods with known methods of using context and domain agents to generate responsive results to users is equivalent to applying a known technique to improve similar products that already operate quite similarly. One of skill in the art would have found it obvious to try any of the teachings of the references in Group 039-B in the systems disclosed in Group 039-A because the teachings of Group 039-B present predictable solutions and skilled artisans would have had a reasonable expectation of success for such combinations. As one example, MMAPS discloses the use of an “agent-based approach” in an “Open Agent Architecture,” where the “agents are distributed entities that can run on different machines, and communicate to solve a task for the user.” Cheyer and Julia, *Multimodal Maps: An Agent-based Approach* (1995) at 3; thus, a person of ordinary skill in the art would be motivated to apply the known methods for using contextually-relevant domain agents, including those disclosed by Group 039-B, and incorporate them into multimodal natural language processing system such as MMAPS. A person of ordinary skill in the art could make a simple substitution of one known element in one reference (e.g., recognition of one form of natural language inputs—speech) for a known element in another (e.g., multimodal natural language input processing) to obtain predictable results. As noted above, commercial incentives may have motivated such a combination.

For similar reasons, it would have been obvious to combine any of the references in Group 039-A with any of the references in Group 039-C, and vice versa. The systems of Group 039-C disclose teachings for natural language processing, which is in the same field of endeavor as the systems of Group 039-A, which likewise involve natural language processing, in addition to the

processing of other inputs. Supplementing the systems of Group 039-A with the teachings of Group 039-C would have amounted to applying one of a finite number of identified, predictable solutions, with a reasonable expectation of success. As one example, MATCH utilized AT&T's Watson Speech Recognition Engine, which made use of "speaker verification" and "speaker dependent models" dependent on speaker-specific information. *See, e.g.*, R. D. Sharp et al., "The Watson speech recognition engine," 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany (1997) at 4067. As such, a person of ordinary skill in the art would be motivated to use known methods for processing natural language inputs in combination with a multimodal natural language processing system such as MATCH.

Finally, in addition to all the reasons stated above, a person of ordinary skill in the art would have been motivated to combine any of the references in Groups 039-A, 039-B, and 039-C with any other reference in any of those groups. All of those references come from the same fields of natural language processing and/or information retrieval. Known work in the field of natural language processing would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference to arrive at the claimed invention, including

for amplifying or improving the processing of user inputs and providing users with tailored and accurate responses.

Additionally, it would have been obvious to combine any of the references in Groups 039-A, 039-B and 039-C with the background references noted above, *see supra* Sections III(A)-(B), for at least the reasons stated above and because the background references all disclose known techniques and methods such as recognizing words and phrases in spoken natural language utterances, using multimodal inputs with speech recognition technologies, accessing information with domain and using user personalization models. It would have been a matter of a simple substitution of these known techniques into the other prior art.

A person of ordinary skill in the art would have been motivated to combine any of the references in Groups 039-A, 039-B, 039-C with any other reference discussed in the Background Section 2(a) above concerning existing speech recognition references. All of those references come from the same fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition would have made use of those references obvious to try and provide a reasonable expectation of success. As inventor Michael Kennewick has previously testified at trial in *VB Assets v. Amazon* matter, speech recognition components were a “commodity product.” Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use

known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

A person of ordinary skill in the art would have been motivated to combine any of the references in Groups 039-A, 039-B, 039-C with any other reference discussed in the Background Section 2(b) above concerning existing prior probability references. All of those references describe use of prior probabilities in the same general field (speech recognition) and/or information retrieval, and the rich history of development in the field of speech recognition generally and the long standing use of prior probability models within that field would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary

skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

A person of ordinary skill in the art would have been motivated to combine any of the references in Groups 039-A, 039-B, 039-C with any other reference discussed in the Background Section 2(c) above concerning existing multimodal references. All of those references describe the use of multiple input modes within the same general fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition and the long standing use of multiple modes of input into systems utilizing speech recognition would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from

one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

A person of ordinary skill in the art would have been motivated to combine any of the references in Groups 039-A, 039-B, 039-C with any other reference discussed in the Background Section 2(d) above concerning existing domain agent references. All of those references describe the use of domain agents in the same fields of speech recognition and/or information retrieval and the rich history of development in the field of speech recognition and the long standing use of domain agents would have made use of those references obvious to try and provide a reasonable expectation of success. Known work in the field of speech recognition would prompt variations of it for use in either the same field or in information retrieval based on design incentives or other market forces because the variations would have been predictable to one of ordinary skill in the art. Moreover, combining prior art elements from one reference with another according to known methods would have yielded predictable results. In addition, it would have been obvious to substitute one known element from a reference for another element from a different reference to obtain predictable results. A person of ordinary skill in the art would have been motivated to use known techniques from one reference to improve similar methods or systems in another reference in the same way. A person of ordinary skill in the art would have been motivated to apply a known technique from one reference to a known method of system from another reference ready for improvement to yield predictable results. A person of ordinary skill in the art would have been

motivated to choose from the finite number of identified, predictable solutions in these references, with a reasonable expectation of success. And many of these references contain a teaching, suggestion, or motivation that would have led one of ordinary skill to modify the prior art reference or to combine it with another reference.

#### **IV. INVALIDITY UNDER 35 U.S.C. § 112**

##### **A. Section 112 ¶ 2: Indefiniteness**

Under pre-AIA Section 112, paragraph 2, a patent’s claims must “particularly point[] out and distinctly claim[] the subject matter which the applicant regards as his invention.” 35 U.S.C. § 112, ¶ 2.<sup>2</sup> As the Supreme Court has explained, “a patent must be precise enough to afford clear notice of what is claimed, thereby apprising the public of what is still open to them.” *Nautilus, Inc. v. Biosig Instruments, Inc.*, 572 U.S. 898, 909 (2014) (cleaned up). “A patent claim is indefinite if it fails to provide a person of ordinary skill in the art reasonable certainty regarding the scope of the claimed invention.” *Synchronoss Techs., Inc. v. Dropbox, Inc.*, 987 F.3d 1358, 1366 (Fed. Cir. 2021); *see also In re Packard*, 751 F.3d 1307, 1310 (Fed. Cir. 2014) (a claim is indefinite “when it contains words or phrases whose meaning is unclear”). In the absence of reasonable certainty about the scope of claims, “competitors cannot avoid infringement, defeating the public notice function of patent claims.” *Halliburton Energy Servs., Inc. v. M-I LLC*, 514 F.3d 1244, 1249 (Fed. Cir. 2008); *see also All Dental Prodx, LLC v. Advantage Dental Prods.*, 309 F.3d 774, 779–80 (Fed. Cir. 2002) (“The primary purpose of the definiteness requirement is to ensure that the claims are written in such a way that they give notice to the public of the extent of the legal protection afforded by the patent, so that interested members of the public, e.g., competitors of the

---

<sup>2</sup> Because the Asserted Patents issued from applications filed on or claiming priority to a date prior to the effective date of the America Invents Act (September 16, 2012), they are subject to the provisions of 35 U.S.C. § 112 predating the AIA amendments. All citations to § 112 herein therefore correspond to the pre-AIA version of the statute.

patent owner, can determine whether or not they infringe.”).

As the Supreme Court explained in *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722 (2002), a patent claim “is a property right; and like any property right, its boundaries should be clear. This clarity is essential to promote progress, because it enables efficient investment in innovation. A patent holder should know what he owns, and the public should know what he does not.” *Id.* at 730–31. Likewise, in *United Carbon Co. v. Binney & Smith Co.*, 317 U.S. 228 (1942), the Supreme Court held that to uphold indefinite patent claims directly contravenes the public interest and discourages innovation:

To sustain claims so indefinite as not to give the notice required by the statute would be in direct contravention of the public interest which Congress therein recognized and sought to protect. . . . The statutory requirement of particularity and distinctness in claims is met only when they clearly distinguish what is claimed from what went before in the art and clearly circumscribe what is foreclosed from future enterprise. A zone of uncertainty which enterprise and experimentation may enter only at the risk of infringement claims would discourage invention only a little less than unequivocal foreclosure of the field.

*Id.* at 233–36.

The Asserted Claims are invalid as indefinite under 35 U.S.C. § 112 ¶ 2, because they fail to inform those of ordinary skill in the art, with reasonable certainty, about the scope of the alleged invention(s). For example, the following claim terms (and their closely related variants) in the Asserted Claims are indefinite, at least under Dialect’s apparent view of the claim scope, rendering invalid all claims containing such terms, as well as their dependent claims:<sup>3</sup>

---

<sup>3</sup> Defendant notes that the attached claim charts identify disclosures in the prior art that correspond to limitations of the Asserted Claims, including in some cases for claim limitations or terms that are identified as indefinite. The identification of disclosures in the prior art for such limitations or terms should not be construed as a concession that Defendant believes those limitations satisfy the definiteness requirement of 35 U.S.C. § 112, ¶ 2, or inform a person of ordinary skill in the art of the full scope of the alleged invention with reasonable certainty. Rather, Defendant’s identification of disclosures in the prior art is based on its best understanding, without the benefit of claim construction, informed by Dialect’s preliminary infringement contentions and in some cases were chosen because they mirror embodiments that the claims of the Asserted Patents purport to cover in the event they are not adjudged to be indefinite.

**'209 patent:** “usable at each stage of processing the received user utterance”; “procedures sensitive to the determined context”; “best responses to the at least one question”; “fuzzy possibilities”; “scraping criteria”; "required values"; "optional values"; "criteria"; "parameters"; "values"; "evaluated values"; "best values."

**'006 patent:** “procedures sensitive to the determined context”; “best responses to the at least one question”; “best values for the one or more required tokens”; “fuzzy possibilities”; “scraping criteria”; "required values"; "optional values"; "criteria"; "parameters"; "further criteria"; "further parameters"; "extracted criteria"; "extracted parameters"; "inferred criteria"; "inferred parameters"; "scoring a relevance."

**'607 patent:** “the device” as recited in claim 12 limitation “the cognitive model includes information on one or more prior interactions between the user and *the device*”; “wherein the cognitive model is a first cognitive model, and wherein the generating the speech-based transcription is based on a second cognitive model that includes information on one or more interactions between the device and a plurality of users of the device.”

**'039 patent:** “textual message”; “searching the query for text combinations”; “a relevance score based on results from comparing the text combinations to entries in the context description grammar”; “obtaining content that is gathered by the selected domain agents”; “wherein the content is arranged in a selected order based on results from the relevance score”; “an aggregate response”; and “fuzzy possibilities.”

**'825 patent:** “fuzzy possibilities”

Defendant’s analysis of the Asserted Patents and Asserted Claims is ongoing. Defendant accordingly reserves the right to supplement this disclosure with additional bases for invalidity based on indefiniteness as that analysis progresses.

## **B. Section 112 ¶ 1: Lack of Written Description**

The Asserted Claims recite limitations that were already well known in the prior art (and therefore are anticipated and obvious). If the claims are construed to recite some novel limitation or combination of limitations, however, then the Asserted Patents fail to describe that limitation or combination.

Section 112 “contains a written description requirement.” *Ariad Pharms., Inc. v. Eli Lilly & Co.*, 598 F.3d 1336, 1351 (Fed. Cir. 2010). The written description requirement requires that “the disclosure of the application relied upon reasonably convey[] to those skilled in the art that the inventor had possession of the claimed subject matter as of the filing date.” *Id.* That is, “the specification must describe an invention understandable to [a] skilled artisan and show that the inventor actually invented the invention claimed.” *Id.*

At least the following claims of the Asserted Patents are invalid for failure to comply with the written description requirement of 35 U.S.C. § 112 ¶ 1, because the specifications of the Asserted Patents fails to reasonably convey to those skilled in the art that the inventors had possession of the claimed subject matter as of the filing date:

**'209 patent:** claims 4, 6, 8, and 9;

**'006 patent:** claims 5–6;

**'607 patent:** claim 12;

**'039 patent:** claims 13–15, 17, and 18; and

As an illustrative example, the '209 patent fails to reasonably convey that the inventors had possession of an invention that included the following limitation(s) as recited in one or more of the Asserted Claims: “matching the keywords contained in the recognized words and phrases to a dynamically updated set of keywords, the dynamically updated set of keywords associated with the dynamic set of prior probabilities or fuzzy possibilities,” “inferring one or more criteria and

one or more parameters using the dynamic set of prior probabilities or fuzzy possibilities,” and “scoring a relevance of each of the asynchronously received results using the dynamic set of prior probabilities or fuzzy possibilities to determine one or more best responses to the at least one question.”

Likewise, as a further illustrative example, the '006 patent fails to reasonably convey that the inventors had possession of an invention that included the following limitation(s) as recited in one or more Asserted Claims: “inferring one or more further criteria and one or more further parameters associated with the request using a dynamic set of prior probabilities or fuzzy possibilities,” “a parser coupled to the processing device,” and “evaluating the one or more extracted values using the dynamic set of prior probabilities or fuzzy possibilities to resolve ambiguous, incomplete, or conflicting information associated with the one or more extracted values.”

As a still further illustrative example, the '607 patent fails to reasonably convey that the inventors had possession of an invention that included the following limitation(s) as recited in one or more Asserted Claims: “generating, by the one or more processors, a non-speech transcription from the non-speech input,” “generating, by the one or more processors, a speech-based transcription based on a cognitive model associated with the user,” “generating, by the one or more processors, a merged transcription from the speech-based transcription and the non-speech transcription,” and “identifying, by the one or more processors, an entry in a context stack, from among a plurality of entries that are in the context stack and that are each indicative of context, wherein the identified entry matches information in the merged transcription.”

As an additional illustrative example, the '039 patent fails to reasonably convey that the inventors had possession of an invention that included the following limitation(s) as recited in one

or more Asserted Claims: “non-speech communications,” merging the speech-based textual message and the non-speech-based textual message,” “searching the query for text combinations,” “comparing the text combinations to entries in a context description grammar,” “accessing a plurality of domain agents,” “generating a relevance score,” “selecting one or more domain agents,” “obtaining content that is gathered by the selected domain agents,” “generating a response/generating an aggregate response,” “generating a context stack,” and “applying prior probabilities or fuzzy possibilities.”

Defendant’s analysis of the Asserted Patents and Asserted Claims is ongoing. Defendant accordingly reserves the right to supplement this disclosure with additional bases for invalidity based on lack of written description as that analysis progresses.

#### **V. PATENT-ELIGIBILITY UNDER 35 U.S.C. § 101**

Section 101 of the Patent Act has long had an important implicit exception, that “[l]aws of nature, natural phenomena, and abstract ideas are not patentable.” *Alice Corp. Pty. Ltd. v. CLS Bank Int’l*, 573 U.S. 208, 216 (2014). In *Mayo Collaborative Servs. v. Prometheus Labs., Inc.* and *Alice Corp. Pty. Ltd. v. CLS Bank Int’l*, the Supreme Court established a two-step framework for applying this exception to Section 101 patent eligibility. *Alice*, 208 U.S. at 217–18; *see also Mayo Collaborative Servs. v. Prometheus Labs., Inc.*, 566 U.S. 66, 71-79 (2012).

The first step is to “determine whether the claims at issue are directed to one of those patent-ineligible concepts”—laws of nature, natural phenomena, or abstract ideas. *Alice*, 208 U.S. at 217; *see Mayo*, 566 U.S. at 77–79. The Federal Circuit, applying the two-step framework for patent-eligibility under *Alice*, has explained that “[r]egardless of what statutory category . . . a claim’s language is crafted to literally invoke,” it is necessary to “look to the underlying invention for patent-eligibility purposes.” *CyberSource Corp. v. Retail Decisions, Inc.*, 654 F.3d 1366, 1374 (Fed. Cir. 2011). In other words, the claims may be “[s]tripped of excess verbiage,” *Intell.*

*Ventures I LLC v. Cap. One Fin. Corp.*, 850 F.3d 1332, 1339 (Fed. Cir. 2017), so that the claimed invention may be viewed in its “simplest form,” *Dealertrack, Inc. v. Huber*, 674 F.3d 1315, 1333 (Fed. Cir. 2012). See also *Intell. Ventures I LLC*, 850 F.3d at 1340 (concluding that claims were, “at their core,” directed to abstract idea); *Affinity Labs of Texas, LLC v. DIRECTV, LLC*, 838 F.3d 1253, 1256 (Fed. Cir. 2016) (“Stripped of excess verbiage, claim 1 is directed to . . . .”); *Elec. Power Grp., LLC v. Alstom S.A.*, 830 F.3d 1350, 1353 (Fed. Cir. 2016) (“[W]e have described the first-stage inquiry as looking at the ‘focus’ of the claims, their ‘character as a whole[.]’”). “[W]ell-established, fundamental concepts fall within the category of abstract ideas.” *Cyberfone Sys., LLC v. CNN Interactive Grp., Inc.*, 558 F. App’x 988, 991 (Fed. Cir. 2014).

If the claims are directed to a patent-ineligible concept, the second step of the patent-eligibility inquiry is to ask “what else is there in the claims”? *Alice*, 573 U.S. at 217; *Mayo*, 566 U.S. at 78. This second step is a search for an “inventive concept,” that is, “an element or combination of elements that is ‘sufficient to ensure that the patent in practice amounts to significantly more than a patent upon the ineligible concept itself.’” *Alice*, 573 U.S. at 217–18; *Mayo*, 566 U.S. at 71–73. “A claim that recites an abstract idea must include ‘additional features’ to ensure ‘that the claim is more than a drafting effort designed to monopolize the abstract idea itself.’” *Alice*, 573 U.S. at 221; see also *Mayo*, 566 U.S. at 77. “‘Simply appending conventional steps, specified at a high level of generality,’ [is] not ‘enough’ to supply an ‘inventive concept.’” *Alice*, 573 U.S. at 222; see also *Mayo*, 566 U.S. at 82. For example, “mere recitation of a generic computer cannot transform a patent-ineligible abstract idea into a patent-eligible invention,” because “[g]iven the ubiquity of computers . . . wholly generic computer implementation is not generally the sort of ‘additional feature’ that provides any ‘practical assurance that the process is more than a drafting effort designed to monopolize the abstract idea itself.’” *Alice*, 573 U.S. at

223–24. In other words, merely adding “purely functional and generic components” to a claim’s recitation of a computer-implemented invention does not “transform an abstract idea into a patent eligible invention” because “[n]early every computer will include [components] capable of performing the basic [] functions required by the [] claims.” *Alice*, 57 U.S. at 226; *see also Two-Way Media Ltd. v Comcast Cable Commc’ns LLC*, 874 F.3d 1329, 1339 (Fed. Cir. 2017) (“Nothing in the claims or their constructions . . . requires anything other than conventional computer and network components operating according to their ordinary functions.”). It is also not enough to “limit[] an abstract idea to one field of use,” or add “token postsolution components,” *Bilski v. Kappos*, 561 U.S. 593, 610–12 (2010), or “limit the use of the idea to a particular technological environment,” *Alice*, 573 U.S. at 223.

As set forth below, the Asserted Claims fail the Supreme Court’s two-step test for patent-eligibility and are therefore patent-ineligible under Section 101. At step one, the Asserted Claims are all directed to abstract ideas, and employ generic, functional, and results-oriented language that recites no specific technological solution for achieving their desired functional results or any improvement in the underlying computer system.

At step two, the Asserted Claims fail to recite any inventive concept sufficient to transform the abstract idea into a patent-eligible invention. Whether viewed individually or as an ordered combination, the elements of the claims recite nothing more than well understood, routine, and conventional facets of computer-implemented speech recognition. To the extent the specifications of the Asserted Patents set forth any kind of implementation details, they do not remedy the generic, functional, results-oriented language of the Asserted Claims. *See Alice*, 573 U.S. at 221 (“[W]e must examine the elements of the claim to determine whether it contains an ‘inventive concept[.]’” (emphasis added)); *RecogniCorp, LLC v. Nintendo Co.*, 855 F.3d 1322, 1327 (Fed.

Cir. 2017) (“To save a patent at step two, an inventive concept must be evident in the claims.”); *Synopsys, Inc. v. Mentor Graphics Corp.*, 839 F.3d 1138, 1149 (Fed. Cir. 2016) (“The § 101 inquiry must focus on the language of the Asserted Claims themselves.”).

Defendant reserves the right to supplement and/or modify these contentions regarding patent-ineligibility, including specifying additional and/or different representative claims, elaborating on the abstract idea(s) to which the Asserted Claims are directed, and further explaining how the Asserted Claims include no more than well understood, routine, and conventional features that do not suffice as an inventive concept at *Alice* step two. Defendant further reserves the right to submit and rely upon expert testimony in support of its contentions regarding patent ineligibility.

#### **A. '209 Patent**

Identification of patent-ineligible claims: Defendant contends that all Asserted Claims of the '209 patent (i.e., claims 1, 4, 6, 7, 8, and 9) are patent-ineligible under Section 101.

Identification of abstract idea: The Asserted Claims of the '209 patent are directed to an abstract idea. For example, representative claim 1 is directed to the abstract idea of using context to execute a spoken request.

Claim elements alleged to be well-understood, routine, and/or conventional: All claim elements are well-understood, routine, and conventional.

Prior art substantiating contention that claim elements are well-understood, routine, and/or conventional: See prior art identified above and in the attached claim charts.

#### **B. '006 Patent**

Identification of patent-ineligible claims: Defendant contends that all Asserted Claims of the '006 patent (i.e., claims 5–6) are patent-ineligible under Section 101.

Identification of abstract idea: The Asserted Claims of the '006 patent are directed to an

abstract idea. For example, representative claim 5 is directed to the abstract idea of using context to execute a spoken request.

Claim elements alleged to be well-understood, routine, and/or conventional: All claim elements are well-understood, routine, and conventional.

Prior art substantiating contention that claim elements are well-understood, routine, and/or conventional: See prior art identified above and in the attached claim charts.

### **C. '607 Patent**

Identification of patent-ineligible claims: Defendant contends that all Asserted Claims of the '607 patent (i.e., claims 12–13) are patent-ineligible under Section 101.

Identification of abstract idea: The Asserted Claims of the '607 patent are directed to an abstract idea. For example, representative claim 12 is directed to the abstract idea of using context and user information to execute a spoken and/or non-verbal request.

Claim elements alleged to be well-understood, routine, and/or conventional: All claim elements are well-understood, routine, and conventional.

Prior art substantiating contention that claim elements are well-understood, routine, and/or conventional: See prior art identified above and in the attached claim charts.

### **D. '039 Patent**

Identification of patent-ineligible claims: Defendant contends that all Asserted Claims of the '039 patent (i.e., claims 13–15, 17, and 18) are patent-ineligible under Section 101.

Identification of abstract idea: The Asserted Claims of the '039 patent are directed to an abstract idea. For example, representative claim 13 is directed to the abstract idea of using context [and user information] to execute a spoken and/or non-verbal request.

Claim elements alleged to be well-understood, routine, and/or conventional: All claim elements are well-understood, routine, and conventional.

Prior art substantiating contention that claim elements are well-understood, routine, and/or conventional: See prior art identified above and in the attached claim charts.

**E. '825 Patent**

Identification of patent-ineligible claims: Defendant contends that all Asserted Claims of the '825 patent (i.e., claims 5–8) are patent-ineligible under Section 101.

Identification of abstract idea: The Asserted Claims of the '825 patent are directed to an abstract idea. For example, representative claim 5 is directed to the abstract idea of using context to execute a spoken request.

Claim elements alleged to be well-understood, routine, and/or conventional: All claim elements are well-understood, routine, and conventional.

Prior art substantiating contention that claim elements are well-understood, routine, and/or conventional: See prior art identified above and in the attached claim charts.