United States District Court,
C.D. California.

PHOENIX SOLUTIONS, INC,
Plaintiff.
v.
The DIRECTV GROUP, INC,
Defendant.

No. CV 08-0984-MRP (SSx)

Nov. 4, 2008.

R. Joseph Trojan, Trojan Law Offices, Beverly Hills, CA, for Plaintiff.

Chris Lind, Jason L. Peltz, Paul J. Skiermont, Bartlit Beck Herman Palenchar and Scott LLP, Chicago, IL, David A. Sergenian, Richard H. Doss, Steven M. Anderson, Quinn Emanuel Urquhart Oliver and Hedges, Los Angeles, CA, for Defendant.

CLAIM CONSTRUCTION ORDER FOR ELEVEN IMPORTANT CLAIM TERMS

MARIANA R. PFAELZER**, District Judge.**

Plaintiff Phoenix Solutions, Inc. ("Phoenix") sues Defendant The DirecTV Group, Inc. ("DirecTV") for alleged infringement of four patents on speech processing technology. Phoenix alleges that DirecTV's operation of a "natural language interactive voice response system" infringes claim 1 of U.S. Patent No. 6,615,172 ("the '172 Patent"); claim 1 of U.S. Patent No. 7,050,977 ("the '977 Patent"); claims 1 and 19 of U.S. Patent No. 7,139,714 ("the '714 Patent"), and claims 1 and 21 of U.S. Patent No. 7,225,125 ("the '125 Patent"). All these patents share substantial portions of the specification and have the same effective filing date.

The Court in this order construes the following eleven terms from the patents in suit:

The '172 Patent:

-> linguistic analysis

-> first linguistic analysis

-> second linguistic analysis

The '977 Patent:

-> client platform

-> signal processing functions

The '714 Patent:

-> client device

-> speech recognition operations

-> first set of speech recognition operations

The '125 Patent:

-> speech processing system

-> distributed speech processing system

-> variants of words to be uttered by users

I.

BACKGROUND

The parties initially proposed numerous disputed claim terms for the Court to construe. Markman v. Westview Instruments, Inc., 517 U.S. 370 (1996). After reviewing the parties' opening claim construction briefs, the Court identified a set of disputed terms that it considered the most important for understanding the asserted patent claims. FN1 The Court and parties held a teleconference on September 24, 2008 where the parties agreed that the Court's set of disputed terms were the most crucial at this point in the litigation . FN2

FN1. The Court originally identified an additional disputed term, "distributed client-server system" from the '125 Patent, for briefing and possible construction at this time. However, the Court declines to construe "distributed client-server system" in this order.

FN2. Defendant DirecTV proposed three more terms for construction and the Court allowed the parties to brief and present oral argument on those terms without agreeing to construe the terms. The terms are: (1) "best match;" (2) "structured language query;" (3) and "webpage." The Court entertained the parties' arguments on those additional terms on the papers and at the hearing. However, the Court declines to construe those terms at this time. The terms are under submission and the Court is prepared to construe them should a motion by either party require their construction.

The parties submitted additional briefing focused on the terms on September 27. The parties on October 27 provided the Court a Technology Tutorial to give it a general background understanding of the relevant technology. *Key Pharm. v. Hercon Laboratories,* 161 F.3d 1576 (Fed.Cir.1996). On the same day, the Court heard oral arguments on the parties' proposed constructions of the disputed terms. With the benefit of these

claim construction proceedings, the Court now defines the identified terms.

II.

LEGAL STANDARD

"The claims of a patent define the invention to which the patentee is entitled" exclusive rights. Phillips v. AWH Corp., 415 F.3d 1303, 1312-13 (Fed.Cir.2005) (en banc) (citations omitted). Construction of the terms of art within the claims is exclusively the province of the Court. Markman v. Westview Instruments, Inc., 517 U.S. 370, 372 (1996). During construction, "[t]he words of a [patent] claim are generally given their ordinary and customary meaning," which is "the meaning that the term would have to a person of ordinary skill in the art in question ... as of the [patent's] effective filing date." Phillips, 415 F.3d at 1313. "Importantly, the person of ordinary skill in the art is deemed to read the claim term not only in the context of the particular claim in which the disputed term appears, but in the context of the entire patent, including the specification." *Id*. The patent's specification is "the single best guide to [its] meaning." *Id*. at 1315.

A court must not, however, import an improper limitation from the specification into a claim by, for example, confining a claim to the embodiments listed in the specification when such a result is not warranted. *Id*. at 1323; E.I. DuPont de Nemours & Co. v. Phillips Petroleum Co., 849 F.2d 1430, 1433 (Fed.Cir.1988). Thus, "[t]he construction that stays true to the claim language and most naturally aligns with the patent's description of the invention [in the specification] will be, in the end, the correct construction." *Phillips* at 1316 (quoting Renishaw PLC v. Marposs Societa' Per Azioni, 158 F.3d 1243, 1250 (Fed.Cir.1998)).

In construing claim terms, a court should also consider the patent's prosecution history, which consists of "the complete record of the proceedings before the PTO [including] the prior art cited during the examination of the patent." *Id*. at 1317. As with the specification, the prosecution history may demonstrate how the examiner and the applicant understood the invention, as "well as whether the inventor limited the invention in the course of prosecution, making the claim scope narrower than it would otherwise be." *Id*.

Finally, a court is permitted to rely on "all evidence external to the patent and prosecution history, including expert and inventor testimony, dictionaries, and learned treatises." *Id*. But this type of evidence is generally "less significant than the intrinsic record in determining the legally operative meaning of claim language." *Id*. (citations omitted).

III.

DISCUSSION

The following tables provide the parties' proposed constructions and Court's adopted constructions.

The '172 Patent			
Term	Court's Construction	Phoenix's Proposed Construction	DirecTV's Proposed Construction
Linguistic analysis	The result of a process that involves parsing symbols to understand, either lexically or	An entire text stream of "recognized words" from the user are	A syntactic analysis of text in order to understand lexically or morphologically the entire linguistic unit (such as determining aspects of the language in the text like noun

	morphologically, a linguistic unit.	analyzed and grouped into larger collections of words, so that an identification of certain "word phrases" is achieved.	phrases), as opposed to word recognition (treating words as symbols with no meaning).
	The concept of "understanding" text must be contrasted with "recognition," which involves resolving digital data into arbitrary symbols rather than extracting some linguistic content or meaning from those symbols.		
First linguistic analysis	The result of a process that involves parsing symbols to understand, either lexically or morphologically, a linguistic unit.	A certain first set of results that underwent "linguistic analysis," as described above.	An initial "linguistic analysis" of the text created by the "continuous speech recognition engine" in order to identify a list of "word phrases" within the text generated by the "speech recognition engine," using analysis of the linguistic characteristics (e.g., part of speech) of each word.
	As used in claim 1, the first linguistic analysis must be the initial linguistic analysis.		
Second linguistic analysis	The result of a process that involves parsing symbols to understand, either lexically or morphologically, a linguistic unit. As used in claim 1, a second linguistic analysis must come after the first linguistic analysis.	A certain second set of results that underwent "linguistic analysis," as described above.	A "linguistic analysis" (which must use results of the "first linguistic analysis") of the text from the database that was obtained as a result of the database query generating the "candidate set of potential matches." The "linguistic analysis" will analyze the text of each database entry returned to determine word phrases present in the text entries of that database that may characterize the content of each such database entry.

The '977 Patent			
Term	Court's Construction	Phoenix's Proposed Construction	DirecTV's Proposed Construction

Client platform	The entire system that makes up the client side of a client-server system, in which at least one device that is part of the platform is physically present where a speech utterance occurs.	A computer system with software adapted for making requests (such as asking for a query to be recognized) to the server computing system.	A combination of hardware and software which permits running client-side software, capable of at least partial speech recognition, operated directly by an end user and equipped with a microphone and a speaker.
Signal processing functions	The speech recognition tasks that convert the digitized speech data into a recognized speech query.	The total set of speech recognition software operations associated with converting the user speech based query into a recognized speech query.	The speech recognition software operations used to convert the digitized speech data into text.

The '714 Patent			
Term	Court's Construction	Phoenix's Proposed Construction	DirecTV's Proposed Construction
Client device	A piece of hardware that is part of the client side of a client-server system. At least one "client device" must be capable of performing some speech recognition operations.	A computing device with software adapted for making requests (such as asking for a query to be recognized) to a network server system.	[same as client platform in '977 Patent:]
			A combination of hardware and software which permits running client-side software, capable of at least partial speech recognition, operated directly by an end user and equipped with a microphone and a speaker.
Speech recognition operations	Processing tasks performed on a digitized speech utterance to recognize symbols contained in the speech utterance.	Processing done by a client device or network server system on speech data to identify words in the utterance.	Computation specifically for the purpose of converting digitized speech to text.
First set of speech recognition operations	A set of speech recognition operations performed on the server side.	The set of speech recognition operations to be performed by the network server system on speech data to identify words in the utterance.	The partial speech recognition processing completed on the "network server system" operating on the "first speech data" being associated with a partial recognition of the speech utterance completed by a "client device."
The '125 Patent			

Term	Court's Construction	Phoenix's Proposed Construction	DirecTV's Proposed Construction
Speech processing system	A system for processing speech-related data.	A system that includes a speech recognition engine or other hardware/software for processing speech data.	A system that includes a "speech recognition engine."
Distributed speech processing system	A speech processing system with components distributed across a network.	A system in which some speech processing operations are performed on a client device which is connected by a network to a separate computing system (such as network server system) where other parts of the speech processing is performed.	A system with different software components at the client's machine client-side and on a server or set of servers, including a client system capable of doing part of the "speech recognition processing" on one end of a network and a server system at the other end of the network capable of doing another part of the "speech recognition processing," where a client is directly operated by the end user and is equipped with a microphone and a speaker.
Variants of words to be uttered by users	Words with the same meaning but that vary by geography, including differences in formation or pronunciation as well as synonyms, so long as the variations are unique to or more probable in a geographic region.	Variations of words or phrases, including a different version, an alternative, modification, or synonym that is used to describe a given word or phrase.	Representative examples of each word from users of different "geographical regions" that correspond to possible differences in a word expected to be seen across a particular language to be recognized, as distinct from word synonyms.

1. The '172 Patent

Phoenix asserts claim 1 of the '172 Patent against DirecTV. The claim requires a "query recognition system" that comprises at least two "linguistic analyses." The patent never expressly defines "linguistic analysis," but the logic of claim 1 and the usage of "linguistic analysis"-and its contrast with "word recognition"-throughout the '172 Patent require the construction that follows. This construction generally comports with the customary usage in the fields of computational linguistics and natural language processing, though "analysis" in the claim at issue refers to the *result* of a process; while in customary usage and the '172 Patent specification "analysis" refers to the *process* itself.

a. Background findings on "word recognition"

The parties do not dispute the term "word recognition." Indeed, the phrase appears throughout the specification but never in the claims. However, the Court considers it necessary to explore the term because the three disputed terms in the '172 Patent are explained in part by contrast to "word recognition."

The first input to the speech query recognition system of claim 1 is an "articulated speech utterance." The parties agree that "articulated speech utterance" means a "vocalized sound presented live by a human that is typically a word or a string of words." Phoenix's Opening Cl. Const. Br., Exh. 2, p. 10. A "continuous speech recognition system" then recognizes words from that speech. '172 Patent cl. 1. The parties agree that a "continuous speech recognition system" is "[a]n automated machine capable ... of identifying ... the words in an 'articulated speech utterance.' " Phoenix's Opening Cl. Const. Brief, Exh. 2, p. 10. The parties' definition is sound and accords with the discussion below, but the concept of "recognition" requires deeper exploration to distinguish "linguistic analysis" from mere word recognition.

"Recognition" in this context refers to a process for translating one set of symbols (here, human utterances) into another set of symbols (machine-readable text). As the '172 Patent explains, one well known form of "recognition" uses the Hidden Markov Model ("HMM"), which can "mathematically describe any time series." '172 Patent col. 3:6-8. The '172 Patent explains that, because "speech is considered to have an underlying sequence of one or more symbols, the HMM models corresponding to each symbol are trained on vectors [derived] from the speech waveforms." Id. col. 3:9-12, 13:9-12. That is, "recognition" is concerned with *describing* something (e.g., sound waves) with a set of *symbols* (e.g., mathematical symbols or variables used by computer programs).

Before recognition can be begin, the analog speech input (wave forms) must be "sampled" (e.g., by taking a measure of the amplitude of the wave at a single point in time) at regular, discrete intervals. This is because analog waveforms are continuous signals-they vary constantly across time-but a digital computer works with discrete values.

Next, the sampled signal is converted to a digital representation of the sample's analog value; that is, a set of digital data that is a nonarbitrary representation of the waves and corresponds to the actual waves. This step is called "quantization." After quantization is complete, the audio is fully converted from analog to a digital data set ready for processing or manipulation by a digital computing system. *See generally,* W.B. KLEIJN & K.K. PALIWAL, SPEECH SYNTHESIS AND CODING, 3-4 (1995). *See also* Hearing Tr. at 50:9-10 (counsel for Phoenix, "Well, certainly, we don't dispute the data has to be digitized"); *infra* n. 11 (explaining how the invention only operates on data that has been digitized).

The system then can try to "recognize" the acoustic features represented by the quantized data. It does so by comparing the quantized data to predefined patterns that correspond to arbitrary symbols (such as digitized text that represents the uttered verbal symbols).FN3 To do the comparison, the system may take the acoustic features from the speech utterance and see if they match up to a set of acoustic features known to the system. Thus, for example, the ' 172 Patent describes an embodiment that "recognizes" speech so that it can be "converted to text." *See also* id. col. 4:20-23 (explaining that natural language processing systems need a "recognizer" that can "recognize" natural speech's "disfluences,-hesitations, repairs and restarts, discourse markets such as 'well' ..." so that these disfluences can be symbolized and annotated by the recognizer "at its input"-that is, before any further operations are performed on the recognized symbols); id. col. 7:19-20 (disclosing another embodiment that "outputs recognized speech text corresponding to the user's question); id. col. 24:60-63 ("By 'recognized' in [the context of one disclosed embodiment] it is meant that the user's query is converted into a text string of distinctive native language words through the HMM technique discussed earlier.").

FN3. The data related to the speech input is nonarbitrary because it is just a mathematical transformation of

the original sampled analog signal. That is, the speech data has a direct mathematical relationship to something that exists in the physical world. The symbols that humans recognize as being embedded in the speech are arbitrary because they do not bear a necessary relationship to the anything in the real world. For example, "dog" in English refers to the animal we know as a dog. But other languages use different symbols to refer to the same animal. Thus, the symbol "dog" is arbitrary, but the data that corresponds to the waveform that contains the word "dog" is not.

The words in the '172 Patent and parties' stipulations-"representing," "identifying," "converting," "describing," and "corresponding"-are alternative ways to capture the essence of recognition as used by the '172 Patent: transforming nonarbitrary digital data that actually corresponds to and represents the sound of a speech utterance (e.g., vectors generated from quantized data) into an arbitrary machine-readable language that represents the symbols contained in the speech utterance (e.g., digitized text).

For these reasons, the Court construes "recognition" to mean "resolving digital data into arbitrary symbols rather than extracting some linguistic content or meaning from those symbols." The modifier "word" indicates that the recognized symbols represent words. Word recognition therefore means "resolving digital data into arbitrary symbols that represent words."

b. Construction of "linguistic analysis"

After generating recognized words from an articulated speech utterance, the system of claim 1 passes the words to a "natural language engine." The engine performs natural language processing (NLP) which involves at least two sets of linguistic analyses on the recognized words. The '172 Patent explains the phrase "linguistic analysis" primarily by contrast to mere word recognition. In the context of illustrating a preferred embodiment, the '172 Patent described an NLP, specifically a "linguistic morphological analysis" that is:

charged with the parsing, understanding and indexing of this large linguistic unit or set of transcribed utterances. The result of this NLP process is to understand lexically or morphologically the entire linguistic unit as opposed to word recognition. Put another way, the linguistic unit or sentence of connected words output by the SRE [speech recognition engine] has to be understood lexically, as opposed to just being "recognized".

Id. col. 12:3-9.

Thus, it is a level of understanding that separates linguistic analysis from word recognition. As the '172 Patent explains: "In contrast to word recognition, Natural language processing (NLP) is concerned with the parsing, understanding and indexing of transcribed utterances and larger linguistic units." '172 Patent at 4:17-18. The applicants agreed. 03/15/2002 Resp. to Office Action, p. 19 ("As mentioned above, linguistic processing is distinct from plain word recognition, and involves concepts such as phrase analysis to fully comprehend an entire closed question posed by a user, not merely picking out words to be used in finding documents.").

Because "understanding" is a slippery concept to apply to a computer, a brief further discussion is warranted. For example, to "understand" the word "dog" in the English text "the friendly dog sits in the room," one must recognize how the word functions in the sentence. To a human English speaker, it is apparent from context that "dog" here is an arbitrary symbol for the animal to which the word refers. *See*

supra n. 3. But before a computer can "understand" what dog means, it must evaluate the text to discover the term's function within the sentence and disambiguate the term. The computer might do this by scanning the string of text to *recognize* the patterns that represent words. This is word recognition. In this example, the computer would recognize "dog" as stem word, an indivisible unit with meaning (to which additional elements such as prefixes, suffixes, or other words to form compounds can be added). Next, the computer can evaluate the stem word "dog" to see how it is employed in the sentence. The computer could determine as which of at least three parts of speech the stem word "dog" is used: (1) as a noun (e.g., the dog ate on the porch); (2) adjective (e.g., her dogged ambition); or (3) verb (e.g., a problem that dogged him for years). Once the word "dog" has some "meaning" to the computer-because it has been associated with the category "noun"-it may be said that the computer has some "understanding" of "dog." *See* '172 Patent col. 19:16-18 (giving another example of the concept); id. col. 34:19-29 (describing a system that embodies such a method). *See also* DANIEL JURAFSKY & JAMES H. MARTIN, SPEECH AND NATURAL LANGUAGE PROCESSING, 57-72 (2000) (illustrating the challenges of such analysis and explaining how to overcome them using a finite-state method such as the HMM).

This understanding lets the computer perform robust operations using the symbol "dog." For instance, without the process, if the computer only "recognized" the symbol "dog," then the computer could do simple tasks such as searching for documents that contains the word; FN4 or the computer could simply present the word on the screen for the user. But, once the hypothetical computer understands that "dog" is a noun, it could be programmed to generate random sentences that use "dog" as the subject of the main verb; or the computer could be programmed to retrieve articles about canines (a text search for "dog") while avoiding short stories about ambitious or troubled people (by running those search results through a word recognizer that filters out all adjectival and verb forms of the root "dog").

FN4. The applicants used this very example. 03/15/2002 Resp. to Office Action, p. 19 ("As mentioned above, linguistic processing is distinct from plain word recognition, and involves concepts such as phrase analysis to fully comprehend an entire closed question posed by a user, *not merely picking out words to be used in finding documents*.") (emphasis added). Similarly, the specification discusses how the root word "driv-," once recognized, can be used to generate other inflectional forms (drives, drove, driving, driven) to generate a more robust search. '172 Patent 19:17-19.

The '172 Patent's specification describes using an understanding of the words in a speech utterance to isolate the most important units in a spoken question. It particularly emphasizes noun phrases (which makes sense in a question, where verb phrases have little content ["How do I ..."; "Where do I ..."; "Can I ..."; "May I ..."] and the noun phrase holds the substance of the question ["How do I pay *with a credit card?* "; "Where do I enter my *credit card* information?"]). *See, e.g.,* '172 Patent col. 7:22-62 (discussing an embodiment that extracts noun phrases from a user's question); col. 37:41-43 (using a credit card example); col. 32:54-67 (parsing a sentence for noun phrases). In a preferred embodiment, the question is then matched to a catalog of answers with the same noun phrase. Id. col. 7:22-62. In such a system, the understanding generated when an NLP runs a *linguistic analysis* on a sentence is then used to *recognize* the parts of an utterance that are important to a task.

The NLP operations described in plain terms above are known as "tokenization" and "tagging" in the art. *See, e.g.,* Jonathan J. Webster, et al., *Tokenization as the Initial Phase in NLP,* COLING-92: Proceedings of the 15th International Conference on Computational Linguistics 1106, 1106, 1109 (1992), *available at* http://www.aclweb.org/anthology-new/C/C92/C92-4173.pdf (last accessed Oct. 1, 2008). Tokens, in almost

all modern linguistic theories, are the smallest meaningful units into which language can be decomposed. *See, e.g.,* Jin Guo, *Critical Tokenization and its Properties,* 23:4 Computational Linguistics 570, 570 (1997), *available at* www.aclweb.org/anthology-new/J/J97/J97-4004.pdf (last accessed Oct. 1, 2008). Tokenization is the process by which the text is resolved into these linguistic units. Id. Tagging, can be seen from the passage below, is the '172 Patent's term for assigning a category marker to a token. *See also* Webster at 1109 (using "tagging" in the same way as the ' 172 Patent).

With these terms of art in mind, it can be seen how the '172 Patent specification illustrates more ways that NLP allows understanding through linguistic analysis:

Tokenization is implemented by a text analyzer which treats the text as a series of tokens or useful meaningful units that are larger than individual characters, but smaller than phrases and sentences. These include words, separable parts of words, and punctuation. Each token is associated with an offset and a length. The first phase of tokenization is the process of segmentation which extracts the individual tokens from the input text and keeps track of the offset where each token originated in the input text. The tokenizer output lists the offset and category for each token. In the next phase of the text analysis, the tagger uses a built-in morphological analyzer to look up each word/token in a phrase or sentence and internally lists all parts of speech. The output is the input string with each token tagged with a parts of speech notation. Finally the grouper which functions as a phrase extractor or phrase analyzer, determines which groups of words form phrases. These three operations which are the foundations for any modern linguistic processing schemes, are fully implemented in optimized algorithms for determining the single-best possible answer to the user's question.

'172 Patent at 17:48-67.

Plaintiff fundamentally misconstrues the nature of "linguistic analysis" by defining it with reference to a possible, but not necessary, final step: analyzing and grouping an entire text stream of recognized words into larger collections of words. Phoenix's Opening Claim Const. Br. 4-5. Plaintiff's proposed construction captures only the last step of some types of linguistic analysis. For example, in the preferred embodiment described above, the analysis takes this final step in the "grouper which functions as a phrase extractor or phrase analyzer, determin[ing] which groups of words form phrases." '172 Patent col. 17:62-64. Plaintiff's construction-ironically, considering Plaintiff's persistent arguments to broaden the scope of the patent-would unduly limit the scope of "linguistic analysis" as well as make the system of claim 1 much more narrow than the claim, properly interpreted, stands.

Thus, the Court concludes that a person of ordinary skill in the arts of computational linguistics and NLP would understand "linguistic analysis," as used in the specification, to refer to the result of the process of understanding, either lexically or morphologically, a linguistic unit. *Cf.* ' 172 Patent col. 12:5-10 (describing how a NLP system might deploy linguistic analysis "to understand lexically or morphologically" linguistic units and contrasting such understanding to "just being 'recognized' ").

However, the '172 Patent itself uses "linguistic analysis" inconsistently, as described above. While the specification, as explored above, uses "linguistic analysis" to refer to a process, the claim in issue uses the term to refer to the result of the analysis process, not the process itself. Further, the claim identifies a "first" and a "second" linguistic analysis; the parties dispute the roles of the first and second analyses. Id. col. 38:48-49, 38:57-59, 38:61-62, 38:64-65.

i. Construction of "first linguistic analysis"

The system of claim 1 performs at least two steps of linguistic analysis on the recognized words. These steps of linguistic analysis must be performed in order because each later step requires an input from the result of a prior step. Thus, "the claim language ... as a matter of logic [requires the steps] be performed in the written order." Altiris, Inc. v. Symantec Corp., 318 F.3d 1363, 1369 (Fed.Cir.2003).

In the first step, "a natural language engine ... generat[es] a first linguistic analysis of said recognized words" and "said first linguistic analysis is used to identify a candidate set of potential matches for said speech utterance." *Id*. col. 38:48-52.

The features of a "first linguistic analysis" in claim 1 that make it more specific than a generic linguistic analysis and any subsequent linguistic analysis are: (1) the "first linguistic analysis" must be the initial linguistic analysis generated because the rest of the system of claim 1 uses the analysis either directly as an input, *id*. col. 38:57-59 (using the first linguistic analysis to generate a candidate set), *id*. col. 38:64-67 (comparing the first linguistic analysis and one or more second linguistic analysis), or uses an input (the candidate set of potential matches) that can only be determined after the first linguistic analysis is returned, *id*. col. 38:62-63 (generating one or more second linguistic analyses from the candidate set); (2) the "first linguistic analysis" is the result of a process because the analysis is " *generat[ed],*" *id*. 38:49 (emphasis added), and then " *used* " to perform later operations, *id*. col. 38:57 (emphasis added); (3) the input the natural language engine uses to generate the analysis are symbols that represent the "recognized words taken from an articulated speech utterance" by the "continuous speech recognition engine," *id*. col. 38:48-51; and (4) the analysis must be "used to identify a candidate set of potential matches." *Id*. col 38:57.

The Court therefore construes "first linguistic analysis" as "the result of a process that involves parsing symbols to understand, either lexically or morphologically, a linguistic unit." As used in each iteration of linguistic analysis in the system of claim 1, there can be only one "first" linguistic analysis. As used in claim 1, input to the first linguistic analysis must come from the "natural language engine." As used in claim 1, the first linguistic analysis is necessary to generate a "candidate set of potential matches."

ii. Construction of "second linguistic analysis"

After the system of claim 1 has identified a candidate set of potential matches using the first linguistic analysis, that candidate set becomes the input for generating "one or more second linguistic analyses." *Id*. at col. 38:51. The system then compares the second linguistic analysis(es) to the first linguistic analysis "to identify a best match" between the "speech utterance" (as represented by the first linguistic analysis) and the candidate set (as represented by the second linguistic analysis(es)). *Id*. at col. 38:64-67.

Therefore, the following features distinguish the second linguistic analysis(es) from the first linguistic analysis: (1) a second linguistic analysis must be generated temporally subsequent to the first linguistic analysis because the input for the second linguistic analysis is the candidate set that can only be generated after the first linguistic analysis, *id*. col. 38:57-59; FN5 and (2) unlike "first linguistic analysis," there may be more than one second linguistic analysis. *Compare id*. col. 38:61, 65 (repeating "one or more second linguistic analyses") *with* 38:48-49, 67, 64 (using an unqualified singular "linguistic analysis"). *See generally* col. 38:44-67 (disclosing an algorithm, each iteration of which can necessarily process only one "first linguistic analysis").

FN5. Plaintiff disregards this requirement. It argues that "[t]hough the first linguistic analysis is compared

with a second linguistic analysis ... that does not impose a requirement that a 'second linguistic analysis' must use the results of a first linguistic analysis.' " Phoenix's Opening Claim Const. Br. 7. Plaintiff is correct that, standing alone, a comparison between a first and second linguistic analysis would not mean that the second linguistic analysis necessarily uses the results of the first. However, Plaintiff fails to account for the fact that the second linguistic analysis, as used in claim 1, cannot be created without the candidate set of potential matches, which requires the first linguistic analysis as an input.

Thus, the Court construes "second linguistic analysis" as "the result of a process that involves parsing symbols to understand, either lexically or morphologically, a linguistic unit." As used in claim 1, there can be multiple second linguistic analyses. As used in claim 1, the input to a second linguistic analysis must come from a "candidate set of potential matches" and therefore the "second linguistic analysis" must come after the process generating the first linguistic analysis.

2. The '977 Patent

Phoenix alleges that DirecTV infringes claim 1 of the '977 Patent. This claim concerns a "speech-enabled internet website." '977 Patent col. 66-67. It appears at this stage of litigation that the central dispute of the parties is the '977 Patent's use of a "client platform" and its allocation of "signal processing functions" between the client platform and a server. Id. col. 39:16-19.

a. Construction of "client platform"

At all times the word "client" is used in the specification, the client is either equated with the person speaking, *see, e.g.,* id. col. 6:61-62 ("[A] client or user can ask a question in a natural language.") or modifies the computing machine physically present where the person speaking is. Id. col. 7:8-9 ("[T]he client's machine is equipped with a microphone and speaker."); id . col. 7:16-19 ("[T]he question that is asked at the client's machine is ... captured by a microphone that is built in as in the case of a notebook computer or is supplied as a standard peripheral attachment."); id. col. 7:17-18 ("The question ... is asked at the client's machine"); id. col. 10:64 ("[A] query articulated by the speaker at the client's machine."); col. 10:65-67 ("[A] speech utterance is captured and partially processed by ... software resident in the client's machine."); id. col. 27:24-25 ("[T]he present invention can accommodate even the worst-case scenario where the client's machine may be quite thin and may have just enough resources to capture the speech input data and do minimal processing.").

Whatever *component* of the "client" *captures* the utterance must physically reside where the speaker is because *every* use of the term in the specification contemplates a client capable of capturing a speech utterance where the person speaking resides. *See* Phillips 415 F.3d at 1317 ("It is ... entirely appropriate for a court, when conducting claim construction, to rely heavily on the written description for guidance as to the meaning of the claims."). The Court here is not imposing an improper or unwarranted limitation from the specification into the claim, but is instead using the term in the only way it can be consistently and naturally read so as to harmonize its usage in both the specification and the claims. Phillips, 415 F.3d at 1316.

This interpretation does not equate the user with the client. Nor does it require that everything that constitutes the "client" or "client side" be physically present at the user's location, where the speech utterance occurs. Such an interpretation would commit the legal error of importing an unwarranted restriction from the specification into the claims. The restriction that is warranted is that some part of the client must be physically present so that it can capture the speech utterance. If some part of the client were

not physically present to capture the speech utterance, the patent would be nonsensical.

By contrast, DirecTV would require every client to be "equipped with a microphone and speaker" by reading a preferred embodiment that uses a microphone and speaker into a limitation on the claim. DirecTV's Opening Claim Const. Br. 10. DirecTV's construction would impose an improper limitation. So long as (1) some part of the client is capable of capturing speech-whether through a "microphone" or any other input device that could capture the soundwaves that make up an articulated speech utterance; and (2) whatever part of the client captures the speech physically resides where the speaker is, then the client should fall within the term's scope.

Nothing in the '977 Patent-and particularly nothing in claim 1-requires that the client be capable of outputting anything by a piece of speaker-like hardware; such an embodiment is mentioned only in the specification's discussion of some preferred embodiments. *See, e.g.*, '977 Patent col. 6:60-7:8-9 ("This interactive system, *when implemented over the World Wide Web* ... [uses a] client's machine [that] is equipped with a microphone and speaker .") (emphasis added).

"Platform" is a very broad term that can include software, hardware, or any combination thereof that provides support for a computer task. *See, e.g.,* DICTIONARY OF COMPUTER AND INTERNET TERMS 380 (Barron's 2003) ("A piece of equipment or software used as a base on which to build something else."); FOLDOC, http://foldoc.org/index.cgi?query=platform (providing several examples of usage). The following quote illustrates a common usage: "The product creates a new platform, the personal computer, on which consumers can play games designed for the Sony PlayStation [platform]." Sony Comp. Enter. v. Connectix Corp., 203 F.3d 596 (9th Cir.2000). "Platform" can be used in narrower senses FN6 but, as explained below, it is the broadest definition that the ' 977 Patent intends.

FN6. *See, e.g.,* FOLDOC (providing several usages of "platform" that relate only to hardware, only to software, or only to a specific hardware-software combination), McGraw-Hill Dict. of Sci. and Tech. (6th ed. 2002) 1615 ("The hardware system and the system software used by a computer program.").

In discussing one embodiment of the invention, the specification notes that "details of [the communications link and a certain process] are not critical, and will vary from platform to platform." '977 Patent col. 27:52-54. Likewise, the specification states:

It will be appreciated by those skilled in the art that the particular implementation for ... processes and routines will vary from client platform to client platform, so that in some environments such processes may be embodied in hard-coded routines executed by a dedicated [digital signal processor], while in others they may be embodied as software routines ... and in still others a combination of the two may be used.

Id. col. 20:10-20.

The passage above requires that the reader of the '977 Patent use the broadest sense of "platform" because the text emphasizes that "platform" refers to an overall system or identifiable collection of hardware, software, or any combination thereof.

The next level up in the '977 Patent's computing ontology is a "device." "Device" refers to a piece of hardware. *See, e.g.,* '977 Patent cl. 18 (referring to "hand held computing devices"). Indeed, the claims of the

'977 Patent suggest that the terms device and platform have different meanings. Id. cl. 19 (using the term "client device" in a claim dependent upon claim 1, which uses only "client platform"); id. cl. 35 (the same). *See also* id. col. 1:45-46 (explaining an improvement over prior art systems that limited users "mostly to non-voice based input/output devices, such as keyboards...."). *See also* Section III.3.a below (construing "client device" and explaining how, because the "client device" need not be the client component that captures the utterance, it need not be physically present at the place of utterance).

To be clear, a client system can be described using the following scheme: (1) at least one device-any identifiable hardware units; (2) at least one platform-a hardware, software, or hardware-software combination that either runs on a device (or system of devices) or is embedded in a device; and (3) processes that run on the platform.FN7

FN7. A system need not have all the layers; nor does this list constitute the entire universe of potential layers. For example, applications can be subdivided into threads, processes, routines, subroutines, functions, etc. *See, e.g.,* col. 20:4-20 (deploying some of these distinctions).

"Client platform" therefore refers to the entire system that makes up the client side of a client-server system.FN8 Though whatever input device receives speech input from the user must be physically present where the speech utterance occurs, nothing in the ' 977 Patent forecloses the possibility that the input device could be physically remote from the remaining components that make up the client platform.

FN8. "Client-side" refers to everything on the opposite side of the network connection from the server. For an illustration, see '977 Patent Fig. 1.

Therefore, "client platform" in the '977 Patent means "the entire system that makes up the client side of a client-server system, in which at least one device that is part of the platform is physically present where a speech utterance occurs."

b. Construction of "signal processing functions"

The parties do not dispute that "signal" refers to a representation of the speech input. Their disputes are (1) whether the signal must be digital; and (2) whether the processing functions convert the signal to "text" or to a "recognized speech query." FN9

FN9. Both parties propose that the signal processing functions must be in "software," but the Court declines to adopt this limitation in its construction. Indeed, both parties disregard the following preferred embodiment disclosed in the specification that explicitly contemplates some of the signal processing be done in hardware or firmware of a dedicated processor: "It will be appreciated by those skilled in the art that the particular implementation for ... processes and routines will vary from client platform to client platform, so that in some environments such processes may be embodied in hard-coded routines executed by a dedicated DSP [digital signal processor]." '977 Patent col. 20:10-20. "DSP" is a common and well established abbreviation for "digital signal processing" or "processor." *See, e.g.,* DICTIONARY OF COMPUTER AND INTERNET TERMS 137 (Barron's 2003); KLEIJN & PALIWAL, 5.

First, Phoenix is correct to note that nothing in the '977 Patent (nor common usage) requires the word "signal," in isolation, to denote a digital signal. *See* Hearing Tr. at 50:9-17. Nowhere does '977 Patent qualify "signal" as analog or digital. However, the signals involved in the invention must take analog form at one or more points (at a minimum, where the speech utterance is received by a piece of hardware such as a microphone; and perhaps again if a signal is output by speakers) and digital at another point (after some degree of processing into computer-manipulable symbols, see Section III. 1.a). However, all the processing functions that the specification and claims discuss can *only* be performed on a digital signal. *See infra* n. 11 (explaining how the mathematical functions that Phoenix argues can be performed on analog data, Phoenix's Reply Cl. Const. Br. 15-16, in fact require digital data). Thus, to construe the '977 Patent's processing functions as capable of operating on analog signals would result in a system that could not function. Cordis Corp. v. Medtronic Ave, Inc., 511 F.3d 1157, 1174 (Fed.Cir.2008) ("To be sure, even outside the means-plus-function [claim construction] context, we have stated that 'a construction that renders the claimed invention inoperable should be viewed with extreme skepticism.' ").

The Court illustrates one of several possible examples of how the language of claim 1 cannot be applied to analog signals to emphasize that the claim uses a recognition routine which can only be performed on digital data. *See supra* Section III. 1.a. In claim 1, the "signal processing functions required to generate said recognized speech query can be allocated between a client platform and the server computing system based on computing resources available," must refer to the processing functions that take place in the speech recognition routine because, on the server side, the "generat[ing of] a recognized speech query" takes place in the speech recognition routine. '977 Patent 39:7-10. It is this recognized speech query-and *only* this recognized speech query-that can be the antecedent of " *said* recognized speech query," *id*. 39:14-15 (emphasis added), because the only previous "recognized speech query" is that in the speech recognition routine. Thus, the signal processing functions are part of the recognition routine. As explained above in Section III.1.a, recognition is done only on digital data.

The Court rejects DirecTV's proposed construction insofar as it would rewrite the claim from "signal processing functions required to generate said recognized speech query" to "signal processing functions required to generate text." As discussed above in Section III. 1.a, the key to language processing is that the signal be converted into symbols that a computer can manipulate; those symbols need not be stored in the computer as "text."

Therefore, the Court concludes that "signal processing functions" are "the speech recognition tasks that convert the digitized speech data into a recognized speech query."

3. The '714 Patent

Claim 1 of the '714 Patent asserted against DirecTV is a method claim directed to a resource allocation scheme that appears similar to one that could be used in the system of claim 1 of the '977 Patent, as discussed above in Section III.2.a. The two patents do not cite one another, however; and the claims of both patents use different terminology ('977 uses "client platform" while '714 uses "client device"; and '977 uses "signal processing functions" while '714 uses "speech recognition operations"). The Court concludes that "client device" differs conceptually from "client platform." On "speech recognition operations," the Court concludes that the operations may include some "signal processing," as discussed in Section III.2.b; and that "speech recognition operations" is broader than "word recognition," as discussed in Section III.1.a.

a. Construction of "client device"

Neither party makes a distinction between "client platform" and "client device." The parties' agreement that "client platform" and "client device" are the same in the claims of these two different patents is understandable, as the distinction is academic as between claim 1 of the '977 Patent and claim 1 of the '714 Patent. Nevertheless, for the sake of completeness, the Court briefly explains the terms' use in '714 Patent. The Court notes that the specifications of the '977 Patent (in which "client platform," but never "client device" appears in the claims) and the '714 Patent (in which "client device, but never "client platform" appears in the claims) make the same distinctions between "device" and "platform." *See supra* Section III.2.a.

As explained above in the '977 Patent construction of "client platform," the terms relate thus: (1) a basic hardware unit or system collectively called the "device"; (2) a hardware, software, or hardware-software combination layer called the "platform" that either runs on the device or is embedded in the device; and (3) processes that run on the platform. Thus, "device" can be both narrower than platform-it can refer to any identifiable hardware unit-and it can be broader than platform-it can refer to a hardware unit that contains several platforms.

The '714 Patent, like the '977 Patent, uses "device" in the specification at all times to refer to an entire physical unit; that is, a piece of hardware may support (or be part of) a platform. The word appears only three times in the specification. '714 Patent col. 1:37 ("The emergence of inexpensive INTERNET [sic] access devices ...."); id. col 1:45-46 (explaining how the invention improves upon the prior art, which was "limited mostly to non-voice based input/output devices, such as keyboards ...."); id. col. 1:62-63 (referring back to the "aforementioned conventional I/O [input/output] devices."). Everywhere else in the specification only "client" or "client platform" is used.

As in the '977 Patent, "device" in the '714 Patent is not precisely synonymous with "platform"; rather, platforms incorporate devices (and perhaps software and firmware). The term "platform" thus refers to entire computing systems that support particular applications or functions. For example, a computer, combined with a microphone and a specific piece of audio processing software could be a "platform" for audio processing; the computer alone can be called a "device"; and the microphone can be called a "device." As applied to this hypothetical, when the '714 Patent speaks of some operations being "completed by a client device," it refers to the hardware portion of the platform known as the "computer" (and perhaps more specifically its central processing unit).

Unlike the system of the '172 Patent, which requires that some component of the client platform be present in the same location as the speech utterance originates because that client platform also captures the speech utterance, *see* Section III.2.a above, nothing in the method of claim 1 in the '714 Patent requires that the "client device" be the actual device that captures the utterance.FN10 Rather, claim 1 of the ' 714 Patent discloses only a method for distributing speech processing between a client and server; it says nothing about how to capture speech. That is, the method of claim 1 of the ' 714 Patent could be used at the point where the system of claim 1 of the ' 977 Patent "allocate[s signal processing functions] between a client platform and the server," '977 Patent col. 39:15-18, but claim 1 of the ' 714 Patent has nothing to say about the device that captures the speech utterance.

FN10. Indeed, this would be impossible in the hypothetical system discussed in the preceding paragraph: the microphone device could not perform the processing; only the computing device could.

At least one client device in the system must be capable of perform partial recognition operations. The '977 Patent was only allowed because the examiner understood it to have "the *limitation* of partial recognition by a client device." 8/17/2006 Reasons for Allowance, p. 2 (emphasis added). Additional reasons why at least one client device must have this capability are discussed below in Section III.3.b.i.

Therefore, "client device" refers to "a piece of hardware that is part of the client side of a client-server system." At least one "client device" must be capable of performing some speech recognition operations.

b. Construction of "speech recognition operations"

The meaning of "recognition" is discussed at length in Section III. 1.a. The modifier "speech" (1) indicates that the recognition operations are performed on articulated utterances that contain meaning, and (2) allows the term to encompass all aspects of recognition operations, of which "word recognition" is one type. *See, e.g.,* Id. 6:6-11 ("A further object of the present invention is to provide a speech recognition system that efficiently integrates a distributed word recognition system with a natural language processing system, so that both individual words and entire speech utterances can be quickly and accurately recognized....").

The parties dispute whether speech recognition operations can be performed on analog input or whether the speech recognition operations can only be performed on digitized speech. As explained in defining "recognition" from the '172 Patent above in Section III.1.a, the recognition step comes after the speech has been quantized; and quantized data is digital data.

The prosecution history of the '714 Patent demands this conclusion. The applicants themselves used the fact that speech recognition can only take place after analog data is converted to digital to overcome a rejection from the PTO. The applicants distinguished a prior art patent where "analog speech is received ... and converted into a stream of digital codes that allow easy transfer on a network." DirecTV Opening Cl. Const. Br. Exh. 8, p. 11. To overcome a PTO rejection that this digital-to-analog conversion (and possible subsequent repackaging into digital data for "easy transfer" with a given network transfer protocol) was a part of speech recognition, the patentee stated that the prior art patent

does not show that the client device assists in any way in any part of the speech recognition process, such as by computing speech vectors, detecting voice activity, performing speech end-pointing, subword/phoneme recognition (such as would be done with Hidden Markov Models and the like), word decoding, etc.

Id. at p. 11.

Notably absent from the patentee's examples of the "speech recognition process" is any operation that has do with converting speech from analog to digital form. More important, sampling and quantizing are prior steps that must be done before any "digital codes" could be transferred on a network. Phoenix will not now be heard to argue that "speech recognition operations," as used in the '172 Patent, can be performed on anything but digital data.FN11

FN11. Phoenix attempts to limit this admission in the prosecution history as responding only to the "conversion of data to a transportable form." Phoenix's Cl. Const. Reply Br. 17. This misses the point: a computer can only work with a digital representation of an analog signal; therefore, digital-to-analog conversion must occur before the computer can perform any operations that convert the "data to a transportable form." Phoenix fundamentally misunderstands the nature of computer networks when it cites

the reference to a dial-up connection in column 21 of the '172 Patent. This part of the specification mentions that the invention might use a "low-speed dial-up connection"; this refers to a modem data connection. Modems (short for "modulator/demodulator") exist precisely to convert (modulate between) digital and analog data: computers use digital data, but phone lines transmit in analog, at least as between the telephone user and a central switching location, Hearing Tr. at 25:23-26:-8, KLEIJN & PALIWAL, 3-5; the place of the modem is to modulate the computer's digital output to an analog signal for transmission across an analog phone line; on the other end of the phone system sits a second modem to demodulate the signal. Similarly, Phoenix cites the discussion in the '714 Patent at cols. 15-16 about how to extract "acoustic features" from speech and misconstrues this as requiring an analog signal. To the contrary, anyone with a cursory understanding of audio processing would understand that the discussion in the cited portion of the specification presupposes digital data. The language from these columns is identical to that discussed in Section III.2.b, above (construing "audio signal processing"). Again, the speech recognition operations are tasks that are a part of the overall signal processing. Just because speech is represented by quantized data does not mean it lacks "acoustic" features. To the contrary, quantized data is a nonarbitrary representation of the actual wave and its acoustic features. *See* Section III. 1.a, above. The discussion Phoenix cites is about performing a cepstral analysis on data that *models* the original analog speech by representing it as "impulse train"; feeds it to a "digital filter"; and performs mathematical transformations on the model. '714 Patent col. 15:29-67, col. 16:1-16. All these things require a digital signal-and it is unassailable that the mathematical transforms discussed in the patent-a discrete cosine transform and fast Fourier transform-require *discrete, i.e., quantized* data points because an analog signal is, by its nature, continuously variable and therefore not discrete. Id.; Section III. 1.a, above. *See also* Wolfram Research Mathworld, *Fast Fourier Transform,* http://mathworld.wolfram.com/FastFourierTransform.html (last accessed Sept. 30, 2008); Stanford University Center for Computer Research in Music and Acoustics, *DFT Defintion,* http://ccrma.stanford.edu/~ jos/mdft/DFT_Definition.html (last accessed Sept. 30, 2008) (explaining that a discrete Fourier transform requires a sampled and quantized signal); DSPRelated.com, *The Discrete Cosine Transform (DCT),* http:// www.dsprelated.com/ds pbooks/mdft/Discrete_Cosine_Transform_DCT.html (last accessed Sept. 30, 2008) (stating and discussing the relationship between FFTs and DCTs). Further, the Patent explains that all the operations are done on a "DSP (or microprocessor)" and use software. Id. col. 16:16-48. DSP is a a common and well established abbreviation for *digital* signal processor or processing. *See, e.g.,* DICTIONARY OF COMPUTER AND INTERNET TERMS 137 (Barron's 2003); KLEIJN & PALIWAL, 5.); and computer software can work only with discrete data points.

The Court therefore construes "speech recognition operations" to be "processing tasks performed on a digitized speech utterance to recognize symbols contain in the speech utterance." The operations required to digitize a speech utterance are not part of the speech recognition operations.

i. "First set of speech recognition operations"

The parties also dispute the meaning of "first set of speech recognition operations," as used in claims 1 and 19 of the '714 Patent. The parties agree the two claims use the terms identically. DirecTV's Opening Cl. Const. Br. Exh. 1, p. 5-6. Claim 1 is directed to a method and claim 19 is directed to a system that implements the method of claim 1.

The first set of speech recognition operations comprises the operations that the are performed by the network server. The contents of the set are determined by "evaluating an amount of computing resources available at the network server system." '714 Patent col. 39:4-6; id. col. 40:9-15 (the same but a system having a

"routine adapted to evaluate" rather than a method for "evaluating"). Both claims 1 and 19 are silent with respect to whether the first set of speech recognition operations contain all the speech recognition operations to be performed or only some of them; it is conceivable that in some implementations of the method, the "evaluat[ion of] an amount of computing resources available at the network server system" could result in a determination that all the speech recognition operations should be performed on the server side. This accords with the specification that anticipates even the "thinnest clients" (i.e., those which have only the most limited processing ability) can be supported. '714 Patent col. 5:17-18.

The thinnest possible client that could be used with the method of claim 1 would have only the ability to convert an analog speech signal to a digital signal and initiate transport of that signal across a network; this hypothetical thinnest client would need to off-load all the speech recognition operations to the server. *See* Section III.3.b, above. Further, claim 1 and claim 19 appear to contemplate the possibility that the first set of speech recognition operations could, in some circumstances, be the only set because the claims say nothing about there any second set of speech recognition operations; the claims do not necessarily require any subsequent set of operations on the server side; and the claims specify no speech recognition operations that must necessarily be done on the client side.

Even if the client does not in fact perform any operations in a particular embodiment, the client must, at a minimum, be capable of partial processing and recognition. The "thinnest" client must be capable of performing some processing because it must, at the absolute minimum, be capable of converting an analog speech signal to digital and transferring the digital signal across a network to the server. *See* Section III.3.b, above.

The specification envisions the scenario where *all* the work is done on the client. '714 Patent col. 22:53-62 ("[F]or some applications, the allocation of signal processing responsibilities may be partitioned differently, to the point where in fact both phases of the speech signal processing may take place at the client side system."). Phoenix illogically argues that this scenario necessarily implies the possibility that a client may not capable, under any circumstances, of performing any operations. Hearing Tr. at 131:11-133:15. Even if this portion of the invention could operate on analog data, this portion of the specification would at most demonstrate that the system can handle the situation where all or nothing should be allocated to one side or the other. But both sides must still have the capability to do some processing or the allocation scheme, the "limitation" that was crux of patentability, would be unnecessary. 8/17/2006 Reasons for Allowance, p. 2.

The Court concludes "first set of speech recognition operations" means "a set of speech recognition operations performed on the server side." The method of claim 1 and system of claim 19 of the '714 allow, but do not require, the first set of speech recognition operations to contain all the speech recognition operations, depending on available resources.

4. The '125 Patent

The '125 Patent claims methods and systems involving distributed speech processing systems that include speech recognition models that are "specifically trained and optimized for a geographic region by using one or more speech models which include variants of words to be uttered by users of the distributed client-server system." '125 Patent abstract. Phoenix asserts claims 1 and 21 of the '125 Patent. The parties agree that the disputed terms have identical meanings in both claims. DirecTV's Opening Cl. Const. Br. 7-8.

a. Construction of "speech processing system"

Speech processing is the broadest of the processing-related terms the Court construes at this time. Again, "speech recognition" refers to the process of recognizing symbols from a speech utterance. *See supra* Section III.1.a. Meanwhile, "signal processing" refers to the processing of a data stream. In the context of the patents at issue, "signal processing" means processing a signal that represents the speech utterance-that is, a signal that corresponds to actual waveforms. *See supra* Section III.2.b. Examples of signal processing functions used in the technology at issue are Discrete Cosine Transforms and Fast Fourier Transforms, which are mathematical transformations of one type of data into another. *See supra* n. 11. Speech recognition therefore requires a signal that has already been processed (e.g., from analog to digital), but signal processing can be done independent of speech recognition (e.g., analog-to-digital conversion with no further steps). Speech processing includes both of the above steps as well as any other processing done on speech-derived or speech-related data.

The parties dispute the bounds of the term speech processing system in the method of claim 1 and the system of claim 21. Phoenix's Reply Cl. Const. Br. 21. Phoenix argues that the term in both claims embraces every possible component of the overall method or system for receiving the initial speech input from the user. Id. 23 (arguing that speech processing system can include "for example, those functions that relate solely to the act of transporting speech data"). Phoenix is correct. The "speech processing" system and method do not address the steps of sampling,FN12 quantizing, or transmitting analog data into the system. Rather, the system and method deal only with the roles of "computing" devices that communicate with one another on a "network server system." Id. cls. 1, 21. The ' 125 Patent has nothing to say about how users interact with the system or how their speech input is received into the system. Instead, the ' 125 Patent only requires that a "speech utterance [come] from a user," but it in no way prescribes how the speech utterance comes from the user into the system.FN13 *See also* id. Fig. 1 (appearing to contemplate that the speech utterance's point of origin rests outside the system).

FN12. The "samples of speech from a group of persons" referred to in the patents is not the same as the process of sampling analog audio for quantization. *See* '125 Patent 39:9. These samples of speech are the data that build the models to which the recognition routines compare the user's speech input to recognize patterns within the speech.

FN13. The '125 Patent does require that the speech samples for building speech recognition models "are communicated [to the system] over a network," but the patent never specifies how the user's speech utterances come into the system. '125 Patent 39:10-11.

The Court therefore concludes that a "speech processing system" is "a system for processing speech-related data."

i. "Distributed speech processing system"

The adjective "distributed" in this context simply refers to components that are physically separate but nevertheless interface with one another. In the context of the claims at issue, the components that make up the processing system are distributed across "a network server system." FN14 *See, e.g., id.* cl. 1, *id.* col. 23:6-10. Note that a network with the feature of being "distributed" need not have a single client side and a single server side. Unlike portions of the specification (and some claims from other patents in suit), see, e.g., ' 125 Patent Fig. 1, which divide a network (or at least the system of the patent) into a binary client-

server world, computers on a distributed network may be "peers"-that is, instead of one serving another, they can both be a "server" and "client" relative to each other. *See* Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd., 545 U.S. 913, 919 (2004). This necessary construction is remarkably broad.FN15

FN14. The Court does not find it necessary to construe "network server system" at this time.

FN15. The prosecution history also demonstrates this term's extraordinary scope. The examiner rejected what became claim 1 of the ' 125 Patent and allowed it only after an amendment to require the method to include "configuring a set of speech recognition operations to be performed by the network server system based on computing resources available to such system." 12/19/2006 Resp. to Final Rejection p. 2. The Statement of Reasons for Allowance explained, "Generally, distributed speech recognition systems having a client/server architecture are well known, but not with *the combined features of optimizing for geographic regions* ***and*** *configuring based upon available resources*." 1/17/2007 Reasons for Allowance p. 2 (emphasis added). The examiner had earlier observed that the only other notable feature, the geographic training aspects of the invention, were obvious. 7/10/06 Nonfinal Rejection p. 7 Therefore, this claim was only allowed because combined three features well known to the prior art: (1) speech processing with (2) configuring based on available resources (i.e., load balancing) and (3) geographic optimization. The same limitation was added to overcome the same reason for rejecting claim 21.12/19/2006 Resp. to Final Rejection p. 6; 1/17/2007 Reason for Allowance p. 3. *See also* 10/05/06 Final Rejection p. 8.

The Court also notes that adding "distributed" to "speech processing system" does not allow devices extraneous to the network server system to become part of the system; rather, "distributed speech processing system" refers to a type of speech processing system and includes all the limitations of "speech processing system" and adds the additional requirement of being "distributed." Additional devices extraneous to the computing devices on the network server system-such an analog system for bringing speech data into a computer on the network system in the first instance-are not part of a distributed speech processing system. As used in claims 1 and 21 of the '125 Patent, a "distributed speech processing system" includes only (1) computing devices where (2) at least some computing devices are capable of performing distributed speech processing and (3) are connected on a network server system.

b. Construction of "variants of words to be uttered by users."

As with the other disputed terms in the '125 Patent, "variants of words to be uttered by users," there is little intrinsic evidence to aid construction. In fact, the term "variants" appears only four times in the entire patent-once in the abstract ("The speech models are trained with samples of word variants expected to be used ... by representative members of a population associated with the geographic region or community of users.") and three times in the claims. '125 Patent at 39:20, 41:16, 42:31. None of these occurrences give the reader much guidance in discerning what constitutes a "variant," putting this patent in tension with its intended notice function. *See* Super. Fireplace Co. v. The Majestic Prods. Co., 270 F.3d 1358, 1371 (Fed .Cir.2001) (discussing the importance of a patent's notice function).

The parties in their papers dispute whether "synonyms" can be "variants" or if only differences in regional pronunciation are "variants." DirecTV in its papers took the relatively untenable position that this capacious word excludes synonyms. *See* DirecTV's Reply Br. at 25. Phoenix's papers, on the other hand, argued that "synonyms" of all sorts are included. *Id*.

At the hearing, both parties moderated their positions and came to the same conclusion that the Court does. Phoenix's counsel stated that "possible variations include synonyms [and] accents [and anything so long as there are] variations across a geographic area," Hearing Tr. at 47:10-15. DirecTV's counsel agreed that "the intrinsic evidence does not support Phoenix's [original construction in the papers] of variance as synonyms divorced from geographic functionality." *Id*. at 122:2-4.

The Court concludes that this is the most natural interpretation of the phrase. In each of the phrase's four occurrences, the variants serve to improve speech recognition so that geographical variations can be processed more efficiently. Therefore, the construction of the term should not be divorced from this goal.

There is no reason why some geographical variations cannot be synonyms as opposed to differences in pronunciation or in a preference for contractions not common in standard written English (e.g., "y'all" vs. "you all"). For example, in California, carbonated beverages are known as "soda." In some parts of Texas, all sodas may be referred to as "Coke." And in the Midwest, sodas are generally "pop." These three words, therefore, are both geographic variants and synonyms. So long as the synonyms vary with reference to a geographic region, they appear to be the type of variations the patent covers.

On the other hand, if the synonyms are not distinct with regard to regional differences in speech, they are not the type of variants the patent contemplates. The definitional question, what constitutes a geographically variable synonym, is answered by the invention's function: if one variation would be more statistically common in a given geographic region than another word that has the same referent, then those words would be geographically variable synonyms. That is, the system's models are built on samples of speech from speakers from different regions. Each model, then, is statistically optimized to best predict the language patterns of a user from a region. ' 125 Patent cols. 13-14.FN16 The ' 125 Patent never expressly says this; instead, as with much else in this Patent, the reader is left to infer it because, as the specification admits, the idea is well known:

FN16. Here, the Court has been careful not to import a limitation (warranted or otherwise) from the specification. The training must generate models based on geographic regions because each independent claim in the ' 125 Patent expressly requires it. '125 Patent 39:8-12 (claim 1), 40:9-13 (claim 13), 41:4-10 (claim 21), 42:19-23 (claim 28).

This training is well-known in the art, so it is not described at length herein, except to note that the distributed architecture of the present invention enhances the quality of HMMs.... In this way, appropriate samples from users of different geographical areas can be easily compiled and analyzed to optimize the possible variations expected....
Id. 14:3-11. *Cf*. Nonfinal Rejection p. 7 (explaining that the geographic training aspects of the invention were obvious).
Therefore, the Court concludes that "variants of words to be uttered by users" means "words with the same meaning but that vary by geography, including differences in formation or pronunciation as well as synonyms, so long as the variations are unique to or more probable in a geographic region."

III.

CONCLUSION

The Court adopts the foregoing constructions of the eleven claim terms identified.

IT IS SO ORDERED.

C.D.Cal.,2008.
Phoenix Solutions, Inc. v. DirecTV Group, Inc.