UNITED STATES PATENT AND TRADEMARK OFFICE
_____

BEFORE THE PATENT TRIAL AND APPEAL BOARD
_____

Ex parte AWNI HANNUN, CARL CASE, JARED CASPER,
BRYAN CATANZARO, GREGORY DIAMOS, ERICH ELSEN,
RYAN PRENGER, SANJEEV SATHEESH,
SHUBHABRATA SENGUPTA, ADAM COATES, and ANDREW Y. NG
_____

Appeal 2018-003323
Application 14/735,002
Technology Center 2600
_____

Before JOHNNY A. KUMAR, JENNIFER L. McKEOWN, and
CATHERINE SHIANG, *Administrative Patent Judges*.

McKEOWN, *Administrative Patent Judge*.

DECISION ON APPEAL

Appellants[1] appeal under 35 U.S.C. § 134(a) from the Examiner's

decision to reject claims 11–20. We have jurisdiction under 35 U.S.C. § 6.

We reverse.

---

[1] According to Appellants, the real party in interest is Baidu USA, LLC.
App. Br. 3.

STATEMENT OF THE CASE

Appellants' disclosed and claimed invention "relates to systems and methods for improving the transcription of speech into text." Spec. ¶ 3. More specifically, Appellants describe that the claimed invention is directed to

> state-of-the-art speech recognition systems developed using end-to-end deep learning. In embodiments, the model architecture is significantly simpler than traditional speech systems, which rely on laboriously engineered processing pipelines; these traditional systems also tend to perform poorly when used in noisy environments. In contrast, embodiments of the system do not need hand-designed components to model background noise, reverberation, or speaker variation, but instead directly learn a function that is robust to such effects. A phoneme dictionary, nor even the concept of a "phoneme," is needed. Embodiments include a well-optimized recurrent neural network (RNN) training system that can use multiple GPUs, as well as a set of novel data synthesis techniques that allows for a large amount of varied data for training to be efficiently obtained. Embodiments of the system can also handle challenging noisy environments better than widely used, state-of-the-art commercial speech systems.

Abstract.

> Claim 11 is illustrative of the claimed invention and reads as follows:

> 11.   A computer-implemented method for transcribing speech comprising:
> receiving an input audio from a user;
> normalizing the input audio to make a total power of the input audio consistent with a set of training samples used to train a trained neural network model;
> generating a jitter set of audio files from the normalized input audio by translating the normalized input audio by one or more time values;
> for each audio file from the jitter set of audio files, which includes the normalized input audio:

generating a set of spectrogram frames for each audio file;
inputting the audio file along with a context of spectrogram frames into a trained neural network;
obtaining predicted character probabilities outputs from the trained neural network; and
decoding a transcription of the input audio using the predicted character probabilities outputs from the trained neural network constrained by a language model that interprets a string of characters from the predicted character probabilities outputs as a word or words.

THE REJECTIONS

The Examiner rejected claims 11–20 under 35 U.S.C. § 101 as directed to patent ineligible subject matter. Final Act. 6–14.

The Examiner rejected claims 11–20 under 35 U.S.C. § 103 as unpatentable over Sompolinsky (US 2011/0035215 A1, published Feb. 10, 2011) and Talwar (2011/0282663 A1, published Nov. 17, 2011). Final Act. 7–14.

ANALYSIS

THE 35 U.S.C. § 101 REJECTION

Claims 11–20

Based on the record before us, we are persuaded that the Examiner erred in rejecting claims 11–20 as directed to patent ineligible subject matter.

An invention is patent-eligible if it claims a "new and useful process, machine, manufacture, or composition of matter." 35 U.S.C. § 101. However, the Supreme Court has long interpreted 35 U.S.C. § 101 to include implicit exceptions: "[l]aws of nature, natural phenomena, and abstract ideas" are not patentable. *Alice Corp. v. CLS Bank Int'l*, 573 U.S. 208, 216 (2014).

In determining whether a claim falls within an excluded category, we are guided by the Supreme Court's two-step framework, described in *Mayo* and *Alice*. *Id.* at 217–18 (citing *Mayo Collaborative Servs. v. Prometheus Labs.*, Inc., 566 U.S. 66, 75–77 (2012)). In accordance with that framework, we first determine what concept the claim is "directed to." *See Alice*, 573 U.S. at 219 ("On their face, the claims before us are drawn to the concept of intermediated settlement, i.e., the use of a third party to mitigate settlement risk."); *see also Bilski v. Kappos*, 561 U.S. 593, 611 (2010) ("Claims 1 and 4 in petitioners' application explain the basic concept of hedging, or protecting against risk.").

Concepts determined to be abstract ideas, and thus patent ineligible, include certain methods of organizing human activity, such as fundamental economic practices (*Alice*, 573 U.S. at 219–20; *Bilski*, 561 U.S. at 611); mathematical formulas (*Parker v. Flook*, 437 U.S. 584, 594–95 (1978)); and mental processes (*Gottschalk v. Benson*, 409 U.S. 63, 69 (1972)). Concepts determined to be patent eligible include physical and chemical processes, such as "molding rubber products" (*Diamond v. Diehr*, 450 U.S. 175, 192 (1981)); "tanning, dyeing, making waterproof cloth, vulcanizing India rubber, smelting ores" (*id.* at 184 n.7 (quoting *Corning v. Burden*, 56 U.S. 252, 267–68 (1854))); and manufacturing flour (*Benson*, 409 U.S. at 69 (citing *Cochrane v. Deener*, 94 U.S. 780, 785 (1876))).

In *Diehr*, the claim at issue recited a mathematical formula, but the Supreme Court held that "[a] claim drawn to subject matter otherwise statutory does not become nonstatutory simply because it uses a mathematical formula." *Diehr*, 450 U.S. at 176; *see also id.* at 192 ("We view respondents' claims as nothing more than a process for molding rubber

products and not as an attempt to patent a mathematical formula."). Having said that, the Supreme Court also indicated that a claim "seeking patent protection for that formula in the abstract . . . is not accorded the protection of our patent laws, . . . and this principle cannot be circumvented by attempting to limit the use of the formula to a particular technological environment." *Id.* (citing *Benson* and *Flook*); *see, e.g., id.* at 187 ("It is now commonplace that an *application* of a law of nature or mathematical formula to a known structure or process may well be deserving of patent protection.").

If the claim is "directed to" an abstract idea, we turn to the second step of the Alice and Mayo framework, where "we must examine the elements of the claim to determine whether it contains an 'inventive concept' sufficient to 'transform' the claimed abstract idea into a patent-eligible application." *Alice*, 573 U.S. at 221 (citation omitted). "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].'" *Id.* (quoting *Mayo*, 566 U.S. at 77). "[M]erely requir[ing] generic computer implementation[] fail[s] to transform that abstract idea into a patent-eligible invention." *Id.*

The PTO recently published revised guidance on the application of section 101. USPTO's January 7, 2019 Memorandum, 2019 *Revised Patent Subject Matter Eligibility Guidance*, 84 Fed. Reg. 50 (Jan. 7, 2019) ("Memorandum"). Under that guidance, we first look to whether the claim recites:

> (1) any judicial exceptions, including certain groupings of abstract
> ideas (i.e., mathematical concepts, certain methods of organizing

human activity such as a fundamental economic practice, or mental

processes); and

(2) additional elements that integrate the judicial exception into a

practical application (see MANUAL OF PATENT EXAMINING PROCEDURE

(MPEP) § 2106.05(a)-(c), (e)-(h) (9th Ed., Rev. 08.2017, Jan. 2018)).

See Memorandum at 52, 55–56. Only if a claim (1) recites a judicial

exception and (2) does not integrate that exception into a practical

application, do we then look to whether the claim:

(3) adds a specific limitation beyond the judicial exception that are not

"well-understood, routine, conventional" in the field (see MPEP

§ 2106.05(d)); or

(4) simply appends well-understood, routine, conventional activities

previously known to the industry, specified at a high level of

generality, to the judicial exception.

See Memorandum at 56.

Examiner's Findings and Conclusion

Under step one of the *Alice* test, the Examiner determines that the

claims are directed to the abstract idea. Final Act. 2–3, 6–7. For example,

the Examiner identifies that the "claimed invention is directed to using the

predicted character probabilities (mathematical formula) to decode a

transcription of the input audio into words or text data." Final Act. 3; *see*

also Final Act. 6–7 (listing the limitations of claim 1 and finding that the

abstract idea "is similar to the court case *Gottschalk v. Benson* because the

predicted character probabilities (mathematical formula or relationship) is

used to convert or transcribe audio data into text data (words)."). The

Examiner determines that the Specification shows that the predicted

character probabilities is an algorithm and, as such, the claimed invention is directed to a mathematical formula. Final Act. 4 (citing Spec. ¶¶ 44, 93).

The Examiner also summarizes the claimed invention into three steps, namely (1) "normalizing the input audio data (manipulating data)", (2) "generating spectrogram frames based on each audio file (generating information sets based on prior information sets)" and (3) "using a mathematical formula to convert audio data into text data (Decoding)." Ans. 4. According to the Examiner, "[m]anipulating data, generating information based on prior information set and Decoding audio data using equations or mathematical formula are all plainly abstract idea category of judicial excepted subject matter" and the abstract ideas are categorized under "'Certain Methods of Organizing Human Activity' since human can listen to an audio file and transcribe the audio data into text data which can all be done mentally." Ans. 4.

Under step two of the *Alice* test, the Examiner determines that the claims do not amount to significantly more than the abstract idea. Final Act. 7. Specifically, the Examiner concludes that the claims "[d]oes not amount to significantly more since it is just decoding a transcription using a mathematical formula or relationship (Predicted Character probabilities). Thus converting or translating audio data into another form of data (text data)." Final Act. 4.

Appellants' Contentions

Appellants, on the other hand, maintain that the Examiner overgeneralizes and oversimplifies the claimed invention and that the claimed invention is not "directed to" an abstract idea. App. Br. 8–9. For example, Appellants assert that the Examiner "tries to eliminate the trained

neural network and related elements" by equating it to a generic computer. According to Appellants,

> A generic computer is not a trained neural network; but even more, a generic computer is not the claimed trained neural network that has been specially designed and trained to receive sets of context of spectrogram frames from a jitter set of audio files, which includes a normalized input audio file obtained from an input audio, to predict character probabilities from the input audio, which are finally selected by being constrained by a language model that interprets a string of characters from the predicted character probabilities outputs as a word or words.

Reply Br. 2.

Appellants also argue that the claimed invention is a specific implementation to address specific technological problems in automatic speech recognition (App. Br. 10–11) and is directed to a "specific improvement in computer capabilities." App. Br. 14; *see also* Reply Br. 2–4 (citing the abstract, paragraphs 93–113, and Table 3 to support that the claimed invention is "an improvement to the technical field of automatic speech recognition."); App. Br. 11–14 (identifying that the claims here, like the claims in *Enfish* and *McRO*, improve the functioning of prior art automatic speech recognition systems).

With respect to step 2 of the *Alice* test, Appellants argue that the claims include significantly more than the abstract idea. Appellants assert that the claims "are not generic processes, such as merely storing and retrieving data," but rather "the claims include specific implementations steps" including the claimed normalizing, generating a jitter set, generating a set of spectrogram frames, obtaining character probabilities, and decoding using the predicted character probabilities. App. Br. 15. Moreover, Appellants assert that the claims include significantly more because

> The claims of the current application deal with improvements to a technology or technical field (e.g., computer functionality via automatic speech recognition, which may be used for numerous purposes including computer interfacing); deal with improvements to the functioning of the computer itself (e.g., by making it easier for users to interface with complex computing systems); deal with a particular machine or system (e.g., computing systems that comprise specifically made models and data); and add specific limitations other than what is well-understood, routine and conventional in the field (e.g., using a specific and non-generic steps to obtain predicted character probabilities outputs and decoding a transcription of the input audio using the predicted character probabilities outputs from the trained neural network constrained by a language model that interprets a string of characters from the predicted character probabilities outputs as a word or words).

App. Br. 16.

Analysis – Revised Step 2A

Under the Memorandum, in prong one of step 2A we look to whether the claim recites a judicial exception. The Examiner identifies the abstract ideas – a mathematical relationship/formula (Final Act. 3) and certain methods of organizing human activity "since human can listen to an audio file and transcribe the audio data into text data which can all be done mentally." Ans. 4.

As an initial matter, we note that the Memorandum identifies mental processes as a separate category of abstract ideas from methods of organizing human activity. We disagree with the Examiner that the claims recite either a method of organizing human activity or a mental process. While transcription generally can be performed by a human, the claims here are directed to a specific implementation including the steps of normalizing an input file, generating a jitter set of audio files, generating a set of

spectrogram frames, obtaining predicted character probabilities from a trained neural network and decoding a transcription of the input audio using the predicted character probability outputs. These are not steps that can practically be performed mentally. Nor do we see how the claimed invention recites organizing human activity. For example, the claims do not include fundamental economic principles or practices, commercial or legal interactions, managing personal behavior or relationships or interactions between people. As such, the claims do not recite a mental process or method of organizing human activity.

The claims do recite using predicted character probabilities to decide a transcription of the input audio, which the Examiner, relying on the Specification, determines is using a mathematical formula. Namely, the Examiner identifies that the Specification discloses an algorithm to obtain the predicted character probabilities. Final Act. 3–4 (citing Spec. 44). The mathematical algorithm or formula, however, is *not* recited in the claims. As such, under the recent Memorandum, the claims do not recite a mathematical concept. *See, e.g.*, *Subject Matter Eligibility Examples: Abstract Ideas*, at 7 (Jan. 7, 2019)(discussing Example 38 and noting that "The claim does not recite a mathematical relationship, formula, or calculation. While some of the limitations may be based on mathematical concepts, the mathematical concepts are not recited in the claims.").

Moreover, even if the claims were considered to recite a mathematical concept, under prong two of step 2A the claims are not directed to an abstract idea because the alleged judicial exception is integrated into a practical application. Namely, as Appellants explain, "the claims of the current application include specific features that were specifically designed

to achieve an improved technological result" and "provide improvements to that technical field." App. Br. 16. For example, the Specification describes that using DeepSpeech learning, i.e. a trained neural network, along with a language model "achieves higher performance than traditional methods on hard speech recognition tasks while also being much simpler." Spec. ¶ 29. As such, based on the record before us, we are persuaded that the Examiner erred in determining that the claims are directed to an abstract idea.

Analysis – Step 2B

We also agree with Appellants that the Examiner fails to sufficiently support the finding that the claims do not add significantly more to the alleged judicial exception. Namely, the Examiner concludes the claims do not include "any additional elements that amounts to significantly more than a judicial exception" but fails to provide sufficient factual support. *See, e.g.*, Final Act. 4, 7; *see also Berkheimer v. HP Inc.*, 881 F.3d 1360 (Fed. Cir. 2018).

Accordingly, we reverse the Examiner's decision to reject claims 11–20 as directed to patent ineligible subject matter.

THE 35 U.S.C. § 103 OBVIOUSNESS REJECTION BASED ON SOMPOLINSKY AND TALWAR

Based on the record before us, we are persuaded that the Examiner erred in rejecting the claims as unpatentable over Sompolinsky and Talwar.

Appellants argue that Sompolinsky fails to teach or suggest decoding a transcription of the input audio using the predicted character probabilities outputs from the trained neural network constrained by a language model that interprets a string of characters from the predicted character

probabilities outputs as a word or words, as required by claims 11 and 17. App. Br. 17. According to Appellants, "[t]he current application employs a model that outputs character-level predictions" whereas Sompolinsky, in contrast, "operates at a word-level or phoneme-level." App. Br 17.

We agree. Sompolinsky describes a speech recognition system that includes neuron models where each neuron model is correlated with a pulse pattern for *a phoneme or entire word utterance*. Sompolinsky ¶¶19–20. In other words, each phoneme or word will have a unique pulse pattern that will be used to identify the phoneme or word. As such, Sompolinsky's system outputs a phoneme, or a phonetic sound, not a character probability. App. Br. 18. We are persuaded of error in the Examiner's determination that Sompolinsky's phoneme output would satisfy the claimed character level probabilities.

We also agree with Appellants that the Examiner erred in relying on Sompolinsky's disclosure of "digits." *See, e.g.*, Ans. 10 (asserting that Sompolinsky discloses transcription models based on digits and digits can be characters). Sompolinsky merely discloses use of the numerical digit words, such as "one" or "eight," as exemplary words for the neuron models. In other words, it is the phoneme output for word, eight, not a digit representation. The Examiner does not sufficiently identify any teaching of the cited combination of Sompolinsky and Talwar that discloses the claimed character level probabilities.

Accordingly, we reverse the Examiner's decision to reject claims 11–20 as unpatentable over Sompolinsky and Talwar.

DECISION

We reverse the Examiner's decision to reject claims 11–20 as directed to patent ineligible subject matter and reverse the Examiner's decision to reject claims 11–20 as unpatentable over Sompolinsky and Talwar.

REVERSED