



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/036,728	09/25/2013	Suzanne Marion Beaumont	RPS920130087USNP(710.257)	8622
58127	7590	09/04/2020	EXAMINER	
FERENCE & ASSOCIATES LLC			ZHU, RICHARD Z	
409 BROAD STREET			ART UNIT	PAPER NUMBER
PITTSBURGH, PA 15143			2675	
			MAIL DATE	DELIVERY MODE
			09/04/2020	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Ex parte SUZANNE MARION BEAUMONT, JAMES ANTHONY HUNT,
ROBERT JAMES KAPINOS, AXEL RAMIREZ FLORES, and
ROD D. WALTERMANN

Appeal 2019-003524
Application 14/036,728
Technology Center 2600

Before ST. JOHN COURTENAY, III, JAMES W. DEJMEK, and
JASON M. REPKO, *Administrative Patent Judges*.

DEJMEK, *Administrative Patent Judge*.

DECISION ON APPEAL

Appellant¹ appeals under 35 U.S.C. § 134(a) from a Final Rejection of claims 1–6, 8–16, 18–20, 23, and 24. Appellant has canceled claims 7, 17, 21, and 22. *See* Appeal Br. 24, 27–28. We have jurisdiction over the remaining pending claims under 35 U.S.C. § 6(b).

We affirm.

¹ Throughout this Decision, we use the word “Appellant” to refer to “applicant” as defined in 37 C.F.R. § 1.42 (2018). Appellant identifies Lenovo (Singapore) PTE. Ltd as the real party in interest. Appeal Br. 3.

STATEMENT OF THE CASE

Introduction

Appellant's disclosed and claimed invention generally relates to identifying a primary speaker "using facial recognition technology in combination with audio analysis." Spec. ¶ 17. In a disclosed embodiment, time-stamped video data indicating visual features associated with speech (e.g., lips moving) is matched with time-stamped audio data to identify a primary speaker from among a group of potential speakers (such as during a video conference). Spec. ¶¶ 32–33. In addition, once a primary speaker has been identified, additional speech content analysis may be performed to separate speech commands from other audio input. Spec. ¶ 41.

Claim 1 is representative of the subject matter on appeal and is reproduced below with the disputed limitations emphasized in *italics*:

1. A method, comprising:

receiving image data from a visual sensor of an information handling device, the image data comprising one or more images of two or more human sources;

receiving audio data from one or more microphones of the information handling device, the audio data comprising speech provided substantially at least partially overlapping from the two or more human sources;

matching the audio data with a pattern of facial features for each of the two or more human sources in the image data;

determining, using the one or more processors and irrespective of a gaze direction from the two or more human sources, a primary speaker from the two or more human sources based on the matching and on identifying that the content of the audio data for one of the two or more human sources is associated with an executable command;

assigning control to the primary speaker based on the determining; and

performing one or more actions based on audio input of the primary speaker.

The Examiner's Rejections

1. Claims 1, 2, 6, 8–12, 16, 18–20, 23, and 24 stand rejected under 35 U.S.C. § 103 as being unpatentable over Li et al. (US 2003/0154084 A1; Aug. 14, 2003) (“Li”); Basson et al. (US 2002/0103649 A1; Aug. 1, 2002) (“Basson”); and Kim et al. (US 2015/0088518 A1; Mar. 26, 2015) (“Kim”). Final Act. 4–10.

2. Claims 3–5 and 13–15 stand rejected under 35 U.S.C. § 103 as being unpatentable over Li, Basson, Kim, and Cloran et al. (US 8,223,944 B2; July 17, 2012) (“Cloran”). Final Act. 10–11.

ANALYSIS²

Appellant argues that the two voice-input commands provided by a user in Kim are not “substantially at least partially overlapping” and are not provided by “two or more human sources,” as recited in claim 1. Appeal Br. 20. Instead, Appellant asserts Kim teaches a single user sequentially provides multiple audible commands. Appeal Br. 20. As such, Appellant argues Kim does not make a determination of a primary speaker because there is only one speaker. Appeal Br. 20. Moreover, Appellant argues Kim fails to teach matching audio data with a pattern of facial features in image data received for each of at least two human sources. Appeal Br. 20–21.

² Throughout this Decision, we have considered the Appeal Brief, filed October 2, 2018 (“Appeal Br.”); the Reply Brief, filed March 25, 2019 (“Reply Br.”); the Examiner’s Answer, mailed January 23, 2019 (“Ans.”); and the Final Office Action, mailed May 2, 2018 (“Final Act.”), from which this Appeal is taken.

Appellant’s arguments are unpersuasive of Examiner error because, at least, they are not responsive to the rejection as articulated by the Examiner. In particular, the Examiner relies on Basson (not Kim) to teach a speaker identification system including the use of microphones to receive audio data in which the audio data comprises “speech provided substantially at least partially overlapping from the two or more human sources.” Final Act. 6 (citing Basson ¶¶ 35, 58, Fig. 2); Ans. 5 (citing Basson ¶¶ 58, 65). Additionally, the Examiner relies on the combined teachings of Li and Basson (not Kim) to teach matching audio data with a pattern of facial features for each of the two or more human sources in the image data to determine a primary speaker. Final Act. 5–6 (citing Li ¶¶ 32, 35, 46, 48, 52–57, 76, 80–81, Figs. 4, 5; Basson ¶¶ 10, 35, 40–41, 58, 62, Fig. 2); Ans. 4–6 (citing Li ¶¶ 9, 70, 76–77; Basson ¶¶ 40–41, 58).

Moreover, the Examiner’s findings are supported by a preponderance of evidence. For example, Basson describes a system comprising a plurality of cameras and microphones to identify a speaker. Basson ¶¶ 2, 9, Fig. 2. Basson describes the cameras are used to capture images of each meeting participant’s face so that a video server may detect which of the participants is speaking “(e.g., based on visually detected lip movement).” Basson ¶¶ 40–41. Basson further describes use of a microphone array using beamforming techniques to switch between alternating or “simultaneous talkers.” Basson ¶ 58. Additionally, Li describes a system for person identification using video-speech matching. Li, Title; *see also* Li ¶¶ 6–10 (describing an embodiment of calculating a correlation between facial images and audio features to determine the speaking person based on the correlation).

Appellant also argues the Examiner failed to support the proposed combination of Li, Basson, and Kim with articulated reasoning as to how and why a person of ordinary skill in the art would have been motivated to combine the references. Appeal Br. 19.

“[T]he law does not require that the references be combined for the reasons contemplated by the inventor.” *In re Beattie*, 974 F.2d 1309, 1312 (Fed. Cir. 1992); see *Outdry Techs. Corp. v. Geox S.p.A.*, 859 F.3d 1364, 1371 (Fed. Cir. 2017). “[A]ny need or problem known in the field of endeavor at the time of invention and addressed by the patent can provide a reason for combining” references. *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 420 (2007). However, we are mindful that although one of ordinary skill in the art may understand that two references *could be* combined as reasoned by the Examiner, this does not imply a motivation to combine the references. *Personal Web Techs., LLC v. Apple, Inc.*, 848 F.3d 987 993–94 (Fed. Cir. 2017); see also *Belden Inc. v. Berk–Tek LLC*, 805 F.3d 1064, 1073 (Fed. Cir. 2015) (“[O]bviousness concerns whether a skilled artisan not only *could have made* but *would have been motivated to make* the combinations or modifications of prior art to arrive at the claimed invention.”); *InTouch Techs., Inc. v. VGO Commc’ns., Inc.*, 751 F.3d 1327, 1352 (Fed. Cir. 2014).

In rejecting claim 1, *inter alia*, the Examiner sets forth reasoning why one of ordinary skill in the art would have been motivated to combine Basson with Li. Final Act. 7. In particular, the Examiner notes that Li explains that as a person speaks, “the person’s head movement causes changes in direction and positions of the face.” Final Act. 7 (citing Li ¶¶ 26, 33). The Examiner finds that an ordinarily skilled artisan would have been motivated to modify Li’s system with the beamforming microphone array of

Basson to determine more readily the location of the speaker for speaker identification and name-face association. Final Act. 7 (citing Basson ¶ 41, Li ¶ 77). Moreover, the Examiner sets forth reasoning why one of ordinary skill in the art would have been motivated to combine Kim's teaching of controlling an electronic device using a voice command with the Li-Basson system. Final Act. 8–9. More specifically, the Examiner reasons it would have been obvious to modify Li's system to identify a primary speaker by determining that content of the received audio data is associated with an executable command (as in Kim) to provide “an alternative way for [a] user to input text and user commands.” Final Act. 8 (citing Li ¶¶ 70, 77). The Examiner further describes the scenario of Li in which different speakers are speaking with varying levels of motion and taking advantage of the established function of Kim to recognize and identify a primary speaker by determining the audio data is associated with an executable command. Final Act. 8–9.

In response to Appellant's arguments, the Examiner further explains why one of ordinary skill in the art would combine Li, Basson, and Kim. Ans. 7–8. Regarding the combination of Li with Basson, the Examiner explains that rather than using an audio segmentation module to separate audio from video obtained by a video camera (as in Li), one of ordinary skill in the art would have been motivated to incorporate the microphone array of Basson to obtain audio-data information to perform the name-face association. Ans. 7; *see also* Final Act. 7. The Examiner explains the proposed combination is the predictable use of prior art elements according to their established functions to obtain a plurality of audio features required by Li. Ans. 7 (citing Li ¶ 9). In addition, the Examiner notes that, similar to

Kim, Li's describes a computer connected to various peripheral devices for, among other things, inputting user commands. Ans. 7–8 (citing Li ¶¶ 70–71, Fig. 2, Kim ¶ 2). Thus, the Examiner determines one of ordinary skill would have been motivated to combine Kim's teaching of determining a primary speaker based on identifying an executable command within a speaker's audio data "to resolve the issue of effectively controlling the information handling device and the connected peripheral devices" as in Li's system. Ans. 8; *see also* Final Act. 7.

Further, the Examiner explains how the proposed combination would work. Ans. 8–11. The Examiner explains that Kim describes two requirements for identifying a speaker as a primary speaker able to control a device. Ans. 8 (citing Kim ¶¶ 51, 59). Kim identifies the requirements as: (i) a speaker must be within the listening range of the device; and (ii) the content of the speaker's command must be associated with a recognized executable command supported by the device. Ans. 8 (citing Kim ¶¶ 51, 58). The Examiner explains the proposed combination uses the established teachings of the modified Li-Basson system (which matches audio data with facial features to determine a speaker's position/location) to determine whether the speaker is within the listening range of the device. Ans. 9 (citing Basson ¶¶ 40–41, 58; Li ¶ 77). In addition, the Examiner explains "then the combination would determine whether the content of the audio data of either one of the two people is associated with an executable command describing attribute information related to the information handling device according to the established functions of *Kim*." Ans. 9 (citing Kim ¶ 59).

Still further, the Examiner finds the predictable use of prior art elements according to their established functions and contributes to a reasonable expectation of success for the proposed combination. Ans. 10–11. The Examiner also finds the proposed combination “any extraordinary skill from one ordinarily skilled in the art.” Ans. 11.

In the Reply Brief, Appellant does not rebut the Examiner’s explanations set forth in the Answer, but instead repeats the arguments made in the Appeal Brief. *See* Reply Br. 20–21.

Here, as detailed above, we find the Examiner has provided the requisite “articulated reasoning with some rational underpinning to support the legal conclusion of obviousness.” *See In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006). Moreover, Appellant has not provided persuasive argument or evidence that the proposed combination uses the elements of the references in a way other than their established functions to achieve predictable results. “The combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results.” *KSR*, 550 U.S. at 416. Further, Appellant does not provide persuasive evidence or reasoning that the proposed combination would be “uniquely challenging or difficult for one of ordinary skill in the art” or “represented an unobvious step over the prior art.” *Leapfrog Enters. Inc. v. Fisher-Price, Inc.*, 485 F.3d 1157, 1162 (Fed. Cir. 2007) (citing *KSR*, 550 U.S. at 418–19).

For the reasons discussed *supra*, we are unpersuaded of Examiner error. Accordingly, we sustain the Examiner’s rejection of independent claim 1. For similar reasons, we also sustain the Examiner’s rejection of independent claims 11, 20, and 24, which recite similar limitations and were

not argued separately. *See* Appeal Br. 18, 21; *see also* 37 C.F.R.

§ 41.37(c)(1)(iv). In addition, we sustain the Examiner’s rejections of claims 2–6, 8–10, 12–16, 18, 19, and 23, which depend directly or indirectly therefrom and were not argued separately. *See* Appeal Br. 18, 21; *see also* 37 C.F.R. § 41.37(c)(1)(iv).

CONCLUSION

We affirm the Examiner’s decision rejecting claims 1–6, 8–16, 18–20, 23, and 24 under 35 U.S.C. § 103.

DECISION SUMMARY

Claims Rejected	35 U.S.C. §	Reference(s)/Basis	Affirmed	Reversed
1, 2, 6, 8–12, 16, 18–20, 23, 24	103	Li, Basson, Kim	1, 2, 6, 8–12, 16, 18–20, 23, 24	
3–5, 13–15	103	Li, Basson, Kim, Cloran	3–5, 13–15	
Overall Outcome			1–6, 8–16, 18–20, 23, 24	

TIME PERIOD FOR RESPONSE

No time period for taking any subsequent action in connection with this appeal may be extended under 37 C.F.R. § 1.136(a). *See* 37 C.F.R. § 41.50(f).

AFFIRMED