UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/243,635	08/22/2016	Rita Singh	0260492	9385

63649   7590   01/30/2019
DISNEY ENTERPRISES, INC.
C/O FARJAMI & FARJAMI LLP
26522 LA ALAMEDA AVENUE, SUITE 360
MISSION VIEJO, CA 92691

EXAMINER
HANG, VU B

ART UNIT	PAPER NUMBER
2672	

NOTIFICATION DATE	DELIVERY MODE
01/30/2019	ELECTRONIC

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.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docketing@farjami.com
farjamidocketing@yahoo.com
ffarjami@farjami.com

UNITED STATES PATENT AND TRADEMARK OFFICE

_____

BEFORE THE PATENT TRIAL AND APPEAL BOARD

_____

Ex parte RITA SINGH and JILL FAIN LEHMAN[1]

_____

Appeal 2018-005675
Application 15/243,635
Technology Center 2600

_____

Before CARLA M. KRIVAK, HUNG H. BUI, and JON M. JURGOVAN, *Administrative Patent Judges.*

KRIVAK, *Administrative Patent Judge.*

DECISION ON APPEAL

Appellants appeal under 35 U.S.C. § 134(a) from the Examiner's Final Rejection of claims 1–20, which are all the claims pending in the application. We have jurisdiction under 35 U.S.C. § 6(b).

We affirm-in-part.

_____

STATEMENT OF THE CASE

Appellants' invention is directed to a method and system "for estimating age of a speaker based on speech" by extracting "a plurality of formant-based feature vectors from each phoneme in the digitized speech based on at least one of a formant position, a formant bandwidth, and a formant dispersion," and comparing the formant-based feature vectors with age determinant formant-based feature vectors (Title (capitalization altered); Abstract).

Claims 1 and 9 are independent. Independent claim 1, reproduced below, is exemplary of the subject matter on appeal.

> 1. A system comprising:
> a microphone configured to receive an input speech from an individual;
> an analog-to-digital (A/D) converter configured to convert the input speech from an analog form to a digital form and generate a digitized speech;
> a memory storing an executable code and an age estimation database including a plurality of age determinant formant-based feature vectors;
> a hardware processor executing the executable code to:
> receive the digitized speech from the A/D converter;
> identify a plurality of boundaries between a plurality of phonemes in the digitized speech;
> extract a plurality of formant-based feature vectors from one or more phonemes of the plurality of phonemes delineated by the plurality of boundaries, based on at least one of a formant position, a formant bandwidth, and a formant dispersion, wherein the formant dispersion is a geometric mean of the formant spacing;
> compare the plurality of formant-based feature vectors with the age determinant formant-based feature vectors of the age estimation database;

estimate the age of the individual when the
comparison finds a match in the age estimation database;
and
communicate an age-appropriate response to the
individual based on the estimated age of the individual.

REJECTIONS and REFERENCES

The Examiner rejected claims 1, 3, 4, 6–9, 11, 12, and 14–20 under
35 U.S.C. § 103 based upon the teachings of Ju (US 2008/0298562 A1,
published Dec. 4, 2008) and Zhao (US 2005/0228664 A1, published Oct. 13,
2005).

The Examiner rejected claims 2, 5, 10, and 13 under 35 U.S.C. § 103
based upon the teachings of Ju, Zhao, and Zigel (US 2015/0351663 A1,
published Dec. 10, 2015).

ANALYSIS

Claims 1–16, 18, and 20

With respect to claim 1, Appellants contend "there is no disclosure,
teaching or suggestion in Zhao about a formant position," where (i)
"[f]ormant position may be *the peak frequency of a formant*" and (ii)
"'formant' means 'the spectral peaks of the sound spectrum |P(f)|' or 'a
concentration of acoustic energy around a particular frequency in the speech
wave'" (App. Br. 9, 12 (citing Spec. 14:5–10); Reply Br. 2). Appellants
further argue Zhao does not teach or suggest "extracting formant-based
feature vectors from phonemes delineated by the boundaries *is based on a
formant position*" as recited in claim 1 (Reply Br. 3–4). We do not agree.

We agree with and adopt the Examiner's findings as our own.
Particularly, we agree with the Examiner *each of* Zhao's (2N+1)*m-

dimension *super vectors 406*, including *acoustic vectors*, teaches *a formant-based feature vector* extracted from phonemes (401, 404 in Zhao's speech waveform 400) delineated by a boundary (402), as required by claim 1 (*see* Zhao ¶¶ 30, 33–35, 46, Fig. 4; Ans. 3; Final Act. 4).

We also agree with the Examiner that Zhao teaches that feature vectors (acoustic vectors in the super vector 406) are extracted from phonemes *based on a formant position*, as recited in claim 1 (Ans. 3). As the Examiner correctly finds, Appellants' Specification does not provide an explicit and exclusive definition of the claimed *formant position*; rather, it broadly describes "a 'formant position' . . . *may be* the peak frequency of a formant" (Ans. 3 (citing Spec. 14:5–10)). Additionally, Appellants' Specification does not define the claimed *feature vectors*, but merely describes that "feature vectors [are extracted] from phonemes of digitized speech 108 *corresponding to one or more formant-based measurements* such as formant positions, formant bandwidth, and/or formant dispersion" (*see* Spec. 7:17–19 (emphasis added)). Based on Appellants' Specification, the Examiner has broadly interpreted the claimed "formant-based feature vectors . . . [extracted] . . . based on . . . a formant position" as encompassing Zhao's *acoustic feature vectors* extracted from a speech waveform at locations where acoustic energy is concentrated (Ans. 3). The Examiner's interpretation and findings are reasonable because Zhao teaches "[t]he *acoustic features can be any of the widely used features such as* MFCCs (Mel Frequency Cepstral Coefficients), LPCs (Linear Prediction Coefficients), LSPs (Line Spectral Pair)/LSF (Line Spectral Frequencies), etc." (*see* Zhao ¶ 35 (emphasis added)). Thus, a skilled artisan, viewing Zhao's teaching, would recognize Zhao's *acoustic features* may be *based on*

spectral peaks in the speech waveform or acoustic energy concentrations at particular locations/frequencies (formant positions), suggesting the claimed extraction of features based on (well known), formant frequencies.[2]

In light of the broad terms recited in claim 1 and the arguments presented, Appellants have failed to clearly distinguish the claimed invention over the prior art relied on by the Examiner. Thus, we sustain the Examiner's rejection of independent claims 1 and 9 argued together and dependent claims 2–8, 10–16, 18, and 20, not separately argued (App. Br. 7, 12–13).

Claims 17 and 19

Dependent claims 17 and 19 recite "the extracting of the plurality of formant-based feature vectors [recited in claim 1 and 9] is based on the formant position, the formant bandwidth, and the formant dispersion." The

---

[2] It was known to the skilled artisan at the time of Appellants' invention that cepstral coefficients, line spectral frequencies (LSFs), and line spectrum pair frequencies (LSPs), may be *based on* spectral formant peaks or acoustic energy concentrations at particular frequencies—*see, e.g.,* Jonathan Lareau, *Application of shifted delta cepstral features for GMM language identification* (Oct. 10, 2006), (unpublished thesis, Rochester Institute of Tech.) (on file with the Thesis/Dissertation Collections at RIT Scholar Works) ("an approximation procedure is used to first find the formant envelope before deriving the Cepstral coefficients"), p. 30, https://scholarworks.rit.edu/cgi/viewcontent.cgi?article=1260&context=thes es; K.K. Paliwal, *On the Use of Line Spectral Frequency Parameters for Speech Recognition*, 2 Digital Signal Processing 80–87 (1992) (describing LSFs based on formant frequencies at page 81); and Robert W. Morris & Mark A. Clements, *Modification of Formants in the Line Spectrum Domain*, 9 IEEE SIGNAL PROCESSING LETTERS, 19, 19–21 (2002) ("we propose an algorithm that takes advantage of the nearly linear relationship between the LSPs and formants").

Examiner finds Zhao's Figure 4 and paragraph 35 teaches the limitations of claim 17 (Final Act. 6).

Appellants argue Zhao does not teach or suggest extraction of formant-based feature vectors based on *all* of the recited *factors including* "the formant bandwidth, **_and_** the formant dispersion" (Reply Br. 5; App. Br. 12–13). We agree with Appellants. Although Zhao's extracted acoustic features "can be any of the widely used features such as MFCCs (Mel Frequency Cepstral Coefficients), LPCs (Linear Prediction Coefficients), LSPs (Line Spectral Pair)/LSF (Line Spectral Frequencies)," Zhao does not teach extracting acoustic features based on both (i) *formant bandwidth* and (ii) *formant dispersion* that is *a geometric mean of the formant spacing,* as recited in claim 1, from which claim 17 depends (*see* Zhao ¶ 35). As the Examiner has not identified sufficient evidence to support the rejection of claim 17, we do not sustain the Examiner's rejection of dependent claims 17 and 19.

DECISION

The Examiner's decision rejecting claims 1–16, 18, and 20 under 35 U.S.C. § 103 is affirmed.

The Examiner's decision rejecting claims 17 and 19 under 35 U.S.C. § 103 is reversed.

No time period for taking any subsequent action in connection with this appeal may be extended under 37 C.F.R. § 1.136(a)(1)(iv).

AFFIRMED-IN-PART