

9UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

KRISP TECHNOLOGIES, INC.,
Petitioner,

v.

SANAS.AI INC.,
Patent Owner.

Case No. PGR2026-00033
Patent No. 12,417,756

PATENT OWNER'S PRELIMINARY RESPONSE

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	THE '756 PATENT	8
III.	CLAIM CONSTRUCTION.....	16
IV.	PETITIONER HAS NOT ESTABLISHED A PRIMA FACIE CASE OF OBVIOUSNESS BECAUSE IT IGNORES THE CLAIMED PIPELINE AND THE EXAMINER-IDENTIFIED INVENTIVE LIMITATION ..	19
	A. The Petition Does Not Demonstrate a Reasoned Motivation to Combine Prior Art References	19
	B. The Petition Does Not Demonstrate a Reasonable Expectation of Success	23
	C. The Petition Does Not Make a Particularized Showing Of Where Each Claim Element Is Disclosed In The Asserted Prior Art.....	26
	D. Institution Should Be Denied as an Inefficient Use of Office Resources.....	27
V.	THE PETITION FAILS TO SHOW THAT ANY CHALLENGED CLAIM IS MORE LIKELY THAN NOT UNPATENTABLE	30
	A. Ground 1 Fails on the Merits.....	31
	1. The Ground 1 Combination is Architecturally Incoherent.....	33
	2. The Nguyen-Dirac Combination Does Not Teach Applying a Trained Machine Learning Model to Extract Accent Features	39
	3. The Proposed Nguyen-Kee-Dirac-Narayanan Combination Does Not Teach Synthesizing the Second User's Speech Based on the First User's Extracted Accent Features.....	41
	4. Nguyen's Architecture Provides No Path for <i>Narayanan's</i> Voice Profile to Inform Synthesis	45
	5. Ground 1 Dependent-Claim Challenges Do Not Cure the Independent-Claim Defects	47
	B. Grounds 2-6 Fail on the Merits	47
	1. Ground 2 Fails Because Kamruzzaman's MFCC Speaker- Identification Features Do Not Teach a Synthesis-Driving Voice Fingerprint	47

2.	Ground 3 Fails Because Wang’s Encoding Step Does Not Provide a Synthesizer That Uses the Encoded Speaker Characteristics	50
3.	Ground 4 Fails Because Nguyen’s Real-Time Media Transport Does Not Teach Real-Time Accent Conversion	53
4.	Ground 5 Fails Because the Petition Merely Repackages the Same Defective Claim 1 and Claim 5 Theories for the Computer-Readable-Medium Claims	55
5.	Ground 6 Fails Because It Merely Aggregates the Same Defective Theories Without a Coherent Accent-Mimicking Architecture.....	56
6.	Grounds 2-6 Fail Because They Rely on Ground 1	56
VI.	CONCLUSION	58

TABLE OF AUTHORITIES

	Page(s)
CASES	
<i>ActiveVideo Networks, Inc. v. Verizon Commc'ns, Inc.</i> , 694 F.3d 1312 (Fed. Cir. 2012)	20
<i>Apple Inc. v. Ginko LLC</i> IPR2025-01388, Paper 14 (PTAB March 2, 2026)	26
<i>Chevron Oronite Co. v. Infineum USA</i> , IPR2018-00923, Paper 9 (PTAB November 7, 2018).....	27
<i>CyberSecure IPS, LLC v. Network Integrity Systems, Inc.</i> , IPR2025-01441, Paper 18 (PTAB May 26, 2026)	27
<i>Honeywell Int'l Inc. v. 3G Licensing, S.A.</i> , 124 F.4th 1345 (Fed. Cir. 2025)	23
<i>Intelligent Bio-Sys., Inc. v. Illumina Cambridge Ltd.</i> , 821 F.3d 1359 (Fed. Cir. 2016)	23
<i>In re Kahn</i> , 441 F.3d 977 (Fed. Cir. 2006)	19, 20, 33
<i>Johns Manville Corp. v. Knauf Insulation, Inc.</i> , IPR2018-00827, Paper 9 (PTAB Oct. 16, 2018)	22
<i>KSR Int'l Co. v. Teleflex Inc.</i> , 550 U.S. 398 (2007).....	19, 20, 33
<i>PAR Pharm., Inc. v. TWI Pharms., Inc.</i> , 773 F.3d 1186 (Fed. Cir. 2014)	23
<i>Xerox Corp. v. Bytemark, Inc.</i> , IPR2022-00624, Paper 9 (PTAB Aug. 24, 2022).....	20
STATUTES	
35 U.S.C. § 103	19
35 U.S.C. § 322(a)(3).....	7, 16, 19, 26

35 U.S.C. § 324(a)	19
37 C.F.R. § 42.204(b)(4).....	26
37 C.F.R. § 42.207	1

PATENT OWNER'S EXHIBITS

<u>EXHIBIT NO.</u>	<u>DESCRIPTION</u>
Ex2001	Complaint filed in <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> , No. 3:25-cv-05666 (N.D. Cal. Jul. 7, 2025) (D.I. 1) (“Complaint”)
Ex2002	Answer filed in <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> , No. 3:25-cv-05666 (N.D. Cal. Sep. 8, 2025) (D.I. 28) (“Answer and Counterclaims”)
Ex2003	Amended Complaint filed in <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> , No. 3:25-cv-05666 (N.D. Cal. Sep. 22, 2025) (D.I. 35) (“Amended Complaint”)
Ex2004	Order on First Motion to Dismiss filed in <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> , No. 3:25-cv-05666 (N.D. Cal. Sep. 29, 2025) (D.I. 37) (“Order on First Motion to Dismiss”)
Ex2005	Order Denying Krisp Motion to Dismiss First Amended Complaint and Granting in Part and Denying in Part Motion to Dismiss Defendant Krisp’s Counterclaims filed in <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> , No. 3:25-cv-05666 (N.D. Cal. Dec. 1, 2025) (D.I. 58) (“Order Denying Krisp Motion to Dismiss First Amended Complaint”)
Ex2006	Answer to Amended Complaint filed in <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> , No. 3:25-cv-05666 (N.D. Cal. Dec. 15, 2025) (D.I. 60) (“Answer to Amended Complaint”)

<u>EXHIBIT NO.</u>	<u>DESCRIPTION</u>
Ex2007	Order Denying Motion to Dismiss filed in <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> , No. 3:25-cv-05666 (N.D. Cal. Dec. 19, 2025) (D.I. 63) (“Order Denying Motion to Dismiss”)
Ex2008	Order Denying [Renewed] Motion for Judgment on Pleadings filed in <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> , No. 3:25-cv-05666 (N.D. Cal. Feb. 23, 2026) (D.I. 72) (“Order Denying [Renewed] Motion for Judgment on Pleadings”)
Ex2009	Joint Case Management Statement filed in <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> , No. 3:25-cv-05666 (N.D. Cal. Mar. 19, 2026) (D.I. 83) (“Joint Case Management Statement”)
Ex2010	Case Management Scheduling Order filed in <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> , No. 3:25-cv-05666 (N.D. Cal. Mar. 26, 2026) (D.I. 92) (“Case Management Scheduling Order”)
Ex2011	Docket Sheet in <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> , No. 3:25-cv-05666 (N.D. Cal.) (“Docket Sheet”)
Ex2012	Coke Morgan Stewart, <i>Memorandum re Interim Processes for PTAB Workload Management</i> (Mar. 26, 2025), https://www.uspto.gov/sites/default/files/documents/InterimProcesses-PTABWorkloadMgmt-20250326.pdf . (“2025 Memo.”)

<u>EXHIBIT NO.</u>	<u>DESCRIPTION</u>
Ex2013	John A. Squires, <i>Memorandum re Additional Discretionary Institution Considerations—U.S. Manufacturing and Small Business Use of AIA Proceedings</i> (Mar. 11, 2026), https://www.uspto.gov/sites/default/files/documents/Additional_Discretionary_Institution_Considerations_US_Manufacturing_and_Small_Business_Use_of_AIA_Proceedings.pdf . (“2026 Memo.”)
Ex2014	<i>Magnolia Med. Techs., Inc. v. Kurin, Inc.</i> , No. IPR2026-00097, Paper 17 (PTAB May 14, 2026) (precedential) (“ <i>Magnolia</i> ”)
Ex2015	Scott R. Boalick, <i>Guidance on USPTO's Rescission of "Interim Procedure for Discretionary Denials in AIA Post-Grant Proceedings with Parallel District Court Litigation"</i> (Mar. 24, 2025)
Ex2016	[Intentionally Left Blank]
Ex2017	Sanas, LinkedIn, https://www.linkedin.com/company/sanastech/about/ (last visited May 13, 2026)
Ex2018	[Intentionally Left Blank]
Ex2019	[Intentionally Left Blank]
Ex2020	[Intentionally Left Blank]
Ex2021	[Intentionally Left Blank]

<u>EXHIBIT NO.</u>	<u>DESCRIPTION</u>
Ex2022	[Intentionally Left Blank]
Ex2023	[Intentionally Left Blank]
Ex2024	<i>About Sanas</i> , Sanas, https://www.sanas.ai/about-us (last visited May 16, 2026).
Ex2025	Krisp, LinkedIn, https://www.linkedin.com/company/krisphq (last visited May 13, 2026)
Ex2026	Krisp Team, <i>What’s It Like to Work at Krisp?</i> (Sep. 20, 2021), https://krisp.ai/blog/working-at-krisp/ (last visited May 13, 2026) (“What’s it like to work at Krisp”)
Ex2027	Arto Minasyan, <i>Krisp Voice AI Lab: Revolutionizing Voice Technology</i> , LinkedIn (posted Mar. 2026), https://www.linkedin.com/posts/artominasyan_eight-years-ago-we-started-with-a-simple-activity-7430219416912629760-8l6t (last visited May 13, 2026).
Ex2028	Krisp, <i>Introducing Accent Conversion for the Listener</i> , https://krisp.ai/blog/introducing-accent-conversion-for-the-listener/ (last visited May 20, 2026)
Ex2029	Arka Telecom, <i>YSU-Krisp AI Lab Opens at Yerevan State University</i> , https://arkatelecom.am/en/news/development/ysu_krisp_ai_lab_opens_at_yerevan_state_university/ (last visited May 20, 2026).

<u>EXHIBIT NO.</u>	<u>DESCRIPTION</u>
Ex2030	Careers at Krisp, Krisp, https://krisp.ai/careers/ (last visited May 14, 2026)
Ex2031	Krisp, Google Maps, https://www.google.com/maps (last visited May 13, 2026)
Ex2032*	Defendant’s Proposed Terms for Construction, <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i>, No. 3:25-cv-05666-RS (KAW) (N.D. Cal. June 9, 2026)
Ex2033*	Defendant’s Invalidity Contentions, <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i>, No. 3:25-cv-05666-RS (KAW) (N.D. Cal. May 26, 2026).

*** New Exhibits submitted herewith.**

I. INTRODUCTION

Pursuant to 37 C.F.R. § 42.207, Patent Owner Sanas.AI Inc. submits this Preliminary Response to the Petition for post-grant review of Claims 1-20 of U.S. Patent No. 12,417,756 (the “’756 Patent”).

Institution should be denied because Petitioner fails to show a coherent system **that combines first-user accent features with second-user natural-voice characteristics to generate output audio that preserves aspects of the second user’s voice while mimicking the first user’s accent**, as required by the claims.

At its core, the Petition is based on impermissible hindsight reconstruction that ignores the very limitation the Examiner identified as supporting allowance: the claimed combination of generating second-user natural-voice characteristics from second input audio and using those generated characteristics with extracted accent features to synthesize speech that preserves aspects of the second user’s voice while mimicking the first accent. Ex1002 at 132. The Petition never identifies any reference—or combination—that discloses or renders obvious that limitation in the context of the claimed pipeline.

Instead, its underlying theory of prior art combinations is illogical. Petitioner cobbles together discrete fragments from four disconnected references directed to different problems and architectures, as illustrated by lynchpin Claim 1:

[1(Pre)] A speech processing system, comprising an **audio interface** coupled to a microphone and an audio output device, memory having

instructions stored thereon, and one or more processors coupled to the memory and the **audio interface** and configured to execute the instructions to:

[1(a)] apply one or more trained machine learning models to first input audio data obtained via the microphone and the **audio interface** to extract accent features of first input speech associated with a first accent of a first user;

[1(b)] analyze obtained second input audio data associated with second input speech associated with a second accent of a second user *to generate characteristics specific to a natural voice of the second user, wherein the generated characteristics correspond to vocal traits that are distinct to the second user and comprise one or more of a voice quality or one or more phonetic patterns, prosodic features, articulation styles, or intonation patterns;*

[1(c)] synthesize a modified version of the second input speech *by modifying the obtained second input audio data based on the generated characteristics* and the extracted accent features, wherein the modified version of the second input speech preserves aspects of the natural voice of the second user and mimics the first accent; and

[1(d)] **provide to the audio interface** output audio data for output via the audio output device, wherein the output audio data is generated based on the modified version of the second input speech.

Ex1001 Cl. 1 (color coding supplied corresponding to the various references as applied by the Petition).

Indeed, the Petition fails to address—let alone map with particularity—the very limitation the Examiner identified in allowing the claims: **[1(b)]** analyzing input data and *generating second-user natural-voice characteristics* and **[1(c)]**

synthesizing *modified second-user speech based on those generated characteristics and the extracted accent features*. See Ex1002 at 132; see also Ex1001 Cls. [7(b)]-[7(c)], [14(b)]-[14(c)].

The Petition sidesteps the Office’s reasoning and seeks to patch together four disparate references in an illogical manner:

- Nguyen (shown in black text) is the Petition’s primary reference, but it discloses a high-level virtual-conference accent-conversion system—**not the claimed real-time accent mimicking pipeline**. Nguyen states that its ML models “accept speech that has one accent and convert it to speech having another accent, while retaining both the content of the original speech and the identity characteristics of the original speaker’s voice.” Ex1004 ¶[0013]. Nguyen likewise describes accent conversion functionality that “receive[s] incoming audio streams that have speech with a source accent and convert[s] it to speech having a target accent, while retaining the identity characteristics of the original speaker.” Ex1004 ¶[0072]. **But those broad statements do not disclose analyzing second-user audio to generate characteristics specific to the second user’s natural voice (as required by [1(b)]) and then synthesizing modified speech based on both those generated characteristics and extracted first-user accent features (as required by [1(c)])**.
- **Kee (shown in blue bold)** is even further removed from the claimed speech-processing pipeline as it relates to “a dual processor computer system operable in a reduced power consumption mode having limited performance.” Ex1012 ¶[0002]. Petitioner relies on **Kee** only for a generic hardware interface: “**An audio interface 162 that is connected to an internal microphone 164 and a pair of speakers 166a, b,**” including “**a digital-to-analog converter**” and “**an analog-to-digital converter.**” Pet. at 16-17. **Kee** says nothing about extracting accent features (as required by [1(a)]), generating natural-voice

characteristics (as required by [1(b)]), or synthesizing modified speech based on both generated voice characteristics and extracted accent features (as required by [1(c)]).

- Dirac (shown in green underline), by contrast, is invoked for accent detection and accent-feature analysis. Dirac teaches that “an audio characteristic analysis of audio communications by the first party may be used to determine the first party’s accent,” and that the audio may be analyzed to obtain characteristics such as “pitch, tone, stress, melody, etc.” Pet. at 23. But Dirac does not teach generating characteristics specific to the natural voice of the second user (as required by [1(b)]), nor using those generated second-user characteristics with extracted first-user accent features to synthesize modified second-user speech that preserves the second user’s natural voice while mimicking the first accent (as required by [1(c)]).
- *Narayanan (shown in purple italics)* is not an accent-conversion reference at all. It is directed to audio correction—identifying incorrect words and replacing them with corrected spoken audio. *Narayanan* explains that audio correction may “*automatically identify any ‘incorrect’ words spoken by a featured speaker*” and correct them “*by replacing them with appropriate ‘correct’ spoken words.*” Ex1006 ¶[0004]. It further teaches generating replacement speech based on a voice profile so the corrected words “resemble or mimic the speaker’s own speech.” Ex1006 ¶[0018]. **That localized edit-and-replace architecture does not teach continuous accent mimicking, much less the claimed synthesis of second-user speech based on generated second-user natural-voice characteristics and extracted first-user accent features (as required by [1(c)]).**

That color-coded patchwork of Nguyen, Kee, Dirac, and *Narayanan* does not constitute a coherent disclosure. At most, it identifies isolated elements drawn from distinct systems that operate on different inputs, employ different architectures, and

yield different outputs. As such, it fails to establish any motivation to combine, reasonable expectation of success, or particularized mapping of claim elements.

This failure is particularly acute with respect to the Examiner-identified limitations, which are disclosed nowhere in the prior art—individually or in any coherent combination.

The Petition does not explain **how** any of the identified elements would be integrated into a single system capable of performing the claimed sequence of operations; **nor** does it **identify** any teaching that aligns those components of generating second-user natural-voice characteristics from second input audio and using those generated characteristics with extracted accent features to synthesize speech that preserves aspects of the second user’s voice while mimicking the first accent as required by [1(b)]. Instead, it purportedly reconstructs the claims from independent references without showing a system that actually performs the claimed method.

This pieced together combination is really an admission that **no single reference—or any coherent combination—teaches the claimed operation. Nor does any reference—or any coherent combination—teach the Examiner-identified requirement of generating second-user natural-voice characteristics from second input audio data and using those generated characteristics, together with extracted accent features, to synthesize modified second-user**

speech that preserves aspects of the second user's natural voice while mimicking the first accent. Each reference **addresses different data, objectives, and architectures,** and the Petition does not explain **how** those fundamentally different systems could be combined (let alone would be combined) to perform the claimed real-time accent-mimicking pipeline, which extracts accent features from first-user speech, generates characteristics specific to the second user's natural voice from second-user speech, and synthesizes modified second-user speech based on both sets of information so that the output preserves the second user's natural voice while mimicking the first user's accent as required by the claims.

Petitioner's theory is essentially taking a high-level virtual-conference accent-conversion framework from Nguyen, borrowing a generic audio interface from [Kee](#), grafting on accent-detection and feature-analysis concepts from [Dirac](#), and importing a same-speaker word-correction voice profile from [Narayanan](#). Having assembled these disparate teachings, Petitioner then declares the result to be the claimed real-time accent-mimicking system. The selected components may all fall under the broad umbrella of speech or audio processing, but they operate on different data, pursue different objectives, and produce different outputs. Their superficial compatibility does not amount to a coherent disclosure or reasoned obviousness showing for the claimed pipeline.

Further, even if cobbled together, the references would not yield the claimed invention, as their combined teachings lack the required arrangement, data flow, and functional interplay for real-time accent mimicking and would at most produce an incompatible patchwork system.

Petitioner **does not identify** any prior-art teaching of the claimed arrangement and operation of a real-time accent-mimicking pipeline—either in a single reference or a coherent combination—it **does not cure** the deficiency identified by the Patent Office. Particularly, it leaves unaddressed the Examiner identified claim elements: generating characteristics specific to the natural voice of the second user from second input audio and using those generated characteristics with extracted accent features to synthesize modified speech that preserves the second user’s natural voice while mimicking the first accent.

This failure is dispositive: the Petition **does not** satisfy its burden under 35 U.S.C. § 322(a)(3) to show **with particularity where each claim element is found in the prior art**, instead reconstructing the claims from disconnected fragments.

Grounds 2-6 fare no better and do not even try to cure Ground 1’s defects. They rely on the same deficient Nguyen-[Dirac](#)-[Narayanan](#)-[Kee](#) core combination and merely add [Kamruzzaman](#), and [Wang](#), where needed for additional claim details. Those additional references **do not supply the missing claim limitations** of generating characteristics specific to the second user’s natural voice, synthesizing a

modified version of the user's speech, and preserving the second user's natural voice.

II. THE '756 PATENT

The '756 Patent is titled "*Systems and methods for real-time accent mimicking.*" The claims address a technical speech-processing problem: modifying speech associated with one accent so that the resulting speech mimics another accent while preserving aspects of the original speaker's natural voice.

The Abstract provides an overview of this invention:

The disclosed technology relates to methods, speech processing systems, and non-transitory computer readable media for real-time accent mimicking. In some examples, trained machine learning model(s) are applied to **first input audio data to extract accent features of first input speech associated with a first accent of a first user**. Obtained second input data associated with second input speech associated with a second accent of a second user is **analyzed to generate characteristics specific to a natural voice of the second user**. A modified version of the second input speech is **synthesized based on the generated characteristics and the extracted accent features**. The modified version of the second input speech advantageously **preserves aspects of the natural voice of the second user and mimics the first accent**. Output audio data generated based on the modified version of the second input speech is provided for output via an audio output device.

Ex1001 Abstract.

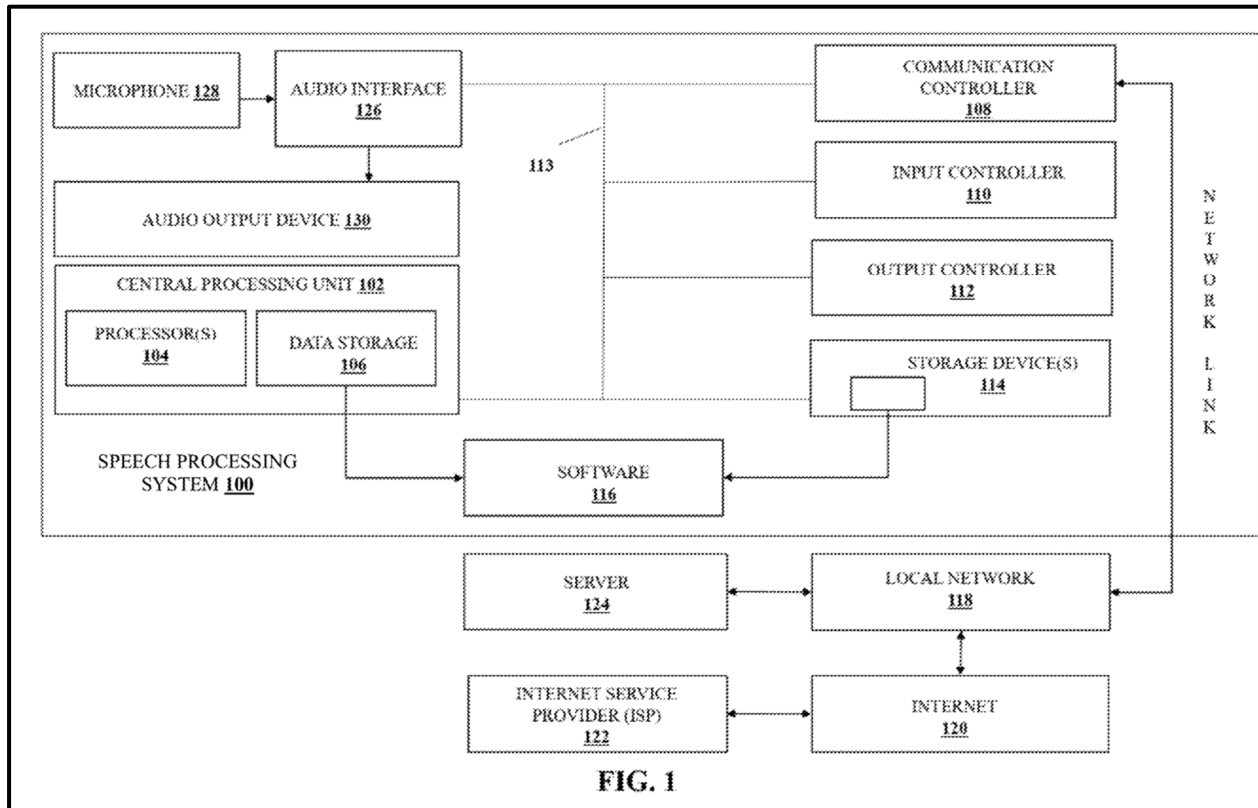
Claim 1 is representative for purposes of this Preliminary Response.¹ It recites a speech processing system with an audio interface, memory, and one or more processors configured to: **[1(a)]** apply one or more trained machine learning models to first input audio data to extract accent features of first input speech associated with a first accent of a first user; **[1(b)]** analyze obtained second input audio data associated with second input speech associated with a second accent of a second user to generate characteristics specific to a natural voice of the second user, where the generated characteristics correspond to vocal traits distinct to the second user; **[1(c)]** synthesize a modified version of the second input speech by modifying the obtained second input audio data based on the generated characteristics and the extracted accent features, where the modified speech preserves aspects of the second user's

¹ Independent Claims 7 and 14 track the same substantive limitations as Claim 1 and confirm that the patent consistently requires the same end-to-end processing pipeline:

- Claim 7 recites a method of the same processing pipeline including the additional limitation of the generated characteristics including a “unique fingerprint of the voice of the second user”. Ex1001 Cl. 7.
- Claim 14 likewise recites a non-transitory computer-readable medium with instructions to perform the same sequence of steps of Claim 1 including the additional limitation of the generated characteristics including “speaker-specific voice characteristics” Ex1001 Cl. 14.

natural voice and mimics the first accent; and [1(d)] provide output audio data for output via an audio output device.

Fig. 1 shows this same architecture in greater detail:

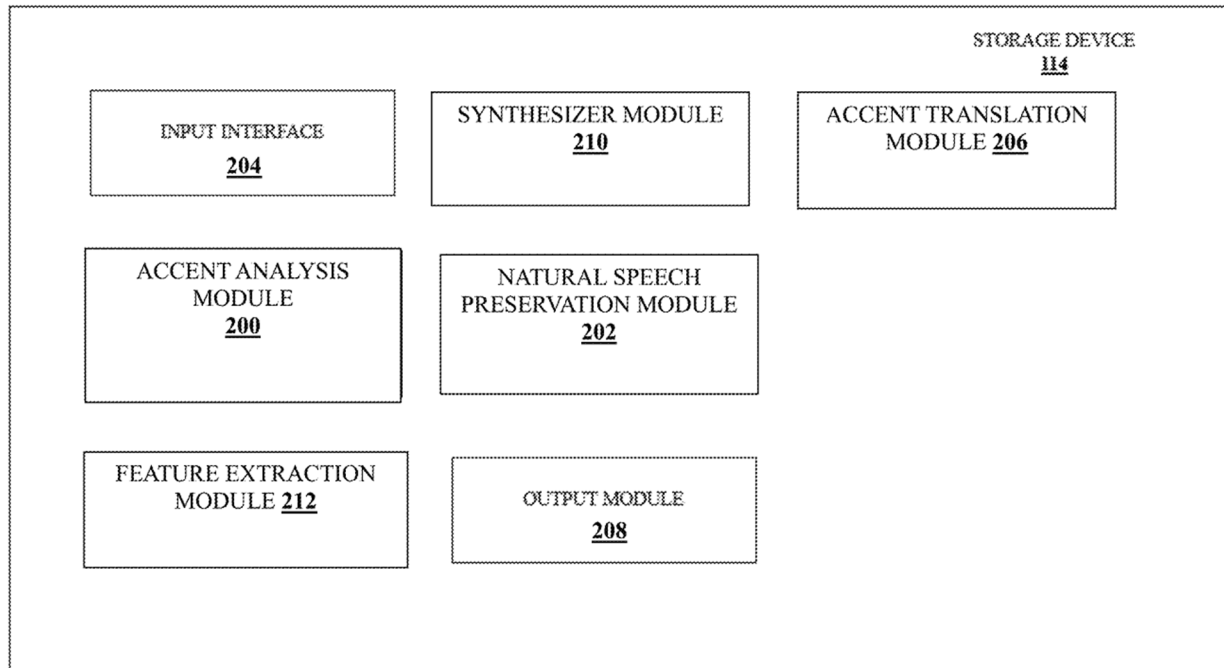


Ex1001 Fig. 1.

As pertinent here, the speech processing system 100 includes, among other components, an audio interface 126, microphone 128, audio output device 130, processor(s) 104, storage device(s) 114, and software 116 executed to perform the real-time accent-mimicking process. Ex1001 Fig. 1. The specification explains that the audio interface 126 “includes sound processing hardware and/or software, including a digital-to-analog converter (DAC) and an analog-to-digital converter

(ADC),” and that it digitizes input audio signals so they can be processed by the speech processing system 100 and later converted for output via the audio output device 130. Ex1001 3:4-19.

Figure 2 illustrates software modules used to perform that processing:



Ex1001 Fig. 2.

In FIG. 2, storage device 114 includes an accent analysis module 200, natural speech preservation module 202, input interface 204, accent translation module 206, output module 208, synthesizer module 210, and feature extraction module 212. Ex1001 Fig. 2, 3:34-42. The specification explains that the input interface 204 receives speech or audio data and “may include hardware components such as

microphones or audio interfaces for capturing real-time speech data,” or software interfaces for prerecorded speech data. Ex1001 3:47-54.

The accent analysis module 200 analyzes first input speech from a first user using machine learning models. “The accent analysis module 200 is configured to analyze input speech from a first user ... using machine learning model(s).” Ex1001 3:65-67. The module “leverages pre-trained machine learning models to analyze captured first input speech and identify accent-specific features,” with models trained on “diverse speech datasets encompassing a wide range of accents.” Ex1001 4:1-7. Those accent features include “pitch contours,” “intonation patterns,” and “phoneme pronunciations,” which “form a critical component for mimicking a first user’s accent in a second user’s speech.” Ex1001 4:8-18.

The feature extraction module 212 likewise extracts relevant speech features. The feature extraction module 212 is configured to extract “linguistic features, prosodic features (e.g., pitch and timbre), and/or global speaker characteristics.” Ex1001 4:19-22. Those global speaker characteristics may include “vocal timbre, speech rate, articulation style, pitch range, rhythm patterns, and/or accent-specific characteristics.” Ex1001 4:22-24. Examples of these characteristics, include timbre as the “unique tonal quality of the speaker’s voice,” speech rate as the speaker’s typical delivery speed, articulation style as the degree of clarity or slurring, pitch

range, rhythm patterns, and accent-specific characteristics such as regional or cultural markers. Ex1001 4:24-43.

The accent translation module 206 translates speech features from the second user's accent to the first user's accent. "The accent translation module 206 is configured to translate the linguistic features extracted by the feature extraction module 212 from a second accent ... to a first accent." Ex1001 4:53-56. The specification explains that the first accent represents "the first user's unique way of pronouncing words and structuring sentences." Ex1001 4:56-58.

The synthesizer module 210 then performs the claimed synthesis by combining the accent-related features with the second user's voice characteristics. "The synthesizer module 210 is configured to combine the extracted accent features, translated linguistic features, extracted prosodic features, and extracted global speaker characteristics to generate a modified version of the second input speech." Ex1001 4:59-63. The natural speech preservation module 202 supports this process by understanding the second user's natural speech characteristics and "substantially maintain[ing] the second user's natural voice during the modification of the second input speech." Ex1001 4:63-67.

The specification further illustrates how the second user's natural voice is preserved. The natural speech preservation module 202 may use "mel frequency cepstral coefficient (MFCC) analysis" to extract "a unique fingerprint of a second

user's voice,” and may also use “speaker identity encoding” to encode “speaker-specific voice characteristics.” Ex1001 5:1-5. These techniques are incorporated into the modified speech so that the system “maintain[s] a natural sound throughout the process of modifying the second input speech” and ensures that the second user's speech “substantially retains its natural quality while mimicking the accent of the first user.” Ex1001 5:8-15.

Thus, the '756 Patent does not merely disclose generic speech processing, accent detection, or voice conversion. It discloses and claims a specific real-time accent-mimicking architecture in which first input speech from a first user is analyzed to extract accent features, second input speech from a second user is separately analyzed to generate characteristics specific to that second user's natural voice, and the system synthesizes a modified version of the second user's speech based on both the extracted first-user accent features and the generated second-user voice characteristics. The resulting output preserves aspects of the second user's natural voice while mimicking the first user's accent. Ex1001, 5:30-6:58; *see also id.* Cl. 1.

The prosecution history underscores the importance of the limitations in the independent claims. The applicant amended the independent claims to require, among other things, analyzing second input audio data to generate characteristics specific to the second user's natural voice and synthesizing modified speech by

modifying the second input audio data based on both the generated characteristics and extracted accent features.

The Examiner then allowed the claims because the prior art failed to teach the claimed combination of:

[1(b)] analyzing obtained second input audio data associated with second input speech associated with a second accent of a second user to generate characteristics specific to a natural voice of the second user, wherein the generated characteristics correspond to vocal traits that are distinct to the second user and comprise one or more of a voice quality or one or more phonetic patterns, prosodic features, articulation styles, or intonation patterns and [[1(c)], [7(c)], and [14(c)]] synthesizing a modified version of the second input speech by modifying the obtained second input audio data based on the generated characteristics and the extracted accent features, [1(c)] wherein the modified version of the second input speech preserves aspects of the natural voice of the second user and mimics the first accent.

Ex1002 at 132 (emphasis and claim elements added).

Yet the Petition *never* substantively addresses this finding. It does **not** *identify any reference* that genuinely discloses this limitation, does **not** sufficiently explain *how* any proposed combination would yield it, and does **not** *reconcile its theory* with the Examiner's contrary conclusion.

Instead, the Petition simply ignores the context of the allowance-dependent limitation while purportedly reconstructing the claims from disparate references—confirming that the Petition **improperly evades the Examiner's reasoning and**

fails to meet its burden under 35 U.S.C. § 322(a)(3) to show where this critical element is disclosed.

III. CLAIM CONSTRUCTION

Petitioner proposes construing “characteristics specific to a natural voice of the second user” as “speaker-specific attributes that are *globally representative* of the second user’s voice and that are used to maintain the second user’s voice throughout synthesis.” Pet. at 5 (emphasis added). That construction contradicts the claim language and should be rejected.

Claim 1 already defines the disputed phrase through its own “wherein” clause: the generated characteristics must “correspond to vocal traits that are distinct to the second user” and “comprise *one or more* of a voice quality or one or more phonetic patterns, prosodic features, articulation styles, or intonation patterns.” Ex1001 Cl. 1 (emphasis added). A single trait satisfies that language yet would not necessarily be “globally representative” of a speaker’s entire voice.

Further, the claims **do not require use “throughout synthesis”**: Element [1(c)] separately requires synthesis “*based on* the generated characteristics and the extracted accent features,” with a result that “preserves *aspects*” of the second user’s natural voice—language materially broader than Petitioner’s continuous-use requirement. Ex1001 Cl. 1 (emphasis added). And where “global” characterization was intended, it was included expressly: Claims 2, 9, and 16 recite extracting “*global*

speaker characteristics”—a distinct term tied to the *first* input audio data—while Claim [1(b)] conspicuously omits any such qualifier. Ex1001 Cls. 2, 9, 16.

In addition, the same claim term (“characteristics specific to a natural voice of the second user”) is currently the subject of claim construction proceedings in the parallel district court litigation between the parties. Ex2032 at 2. Notwithstanding Petitioners’ assertion that their “broad *Sotera* stipulation [] eliminates any overlap between this PGR and the district court litigation,” Petitioner has requested that the district court construe the very same term at the Markman hearing scheduled for October 2026—after the expected institution decision in this case, but before any final written decision would be due. Petitioner’s Response to Patent Owner’s Discretionary Denial Brief, Paper 8 at 6. The district court proceedings will address the scope and meaning of the asserted claims based on a more developed record, including full briefing, discovery, and expert testimony. Proceeding with claim construction in parallel here risks inconsistent interpretations, duplicative efforts, and a decision on an incomplete record, further underscoring that institution would not promote efficient resolution of the dispute.

In any event, for purposes of institution, *it does not matter which construction the Board adopts because Petitioner’s analysis fails either way.* Accordingly, Patent Owner submits that no express construction is necessary at this preliminary stage. To the extent construction is required, the Board should reject

Petitioner's attempt to import exemplary compression and dimensionality-reduction embodiments into the claims. Patent Owner reserves the right to propose express constructions should trial be instituted.

IV. PETITIONER HAS NOT ESTABLISHED A PRIMA FACIE CASE OF OBVIOUSNESS BECAUSE IT IGNORES THE CLAIMED PIPELINE AND THE EXAMINER-IDENTIFIED INVENTIVE LIMITATION

The Board may deny institution where a petitioner fails to satisfy its threshold burdens, which in the present case includes demonstrating:

- (1) a reasoned motivation to combine prior art references,
- (2) a reasonable expectation of success, and
- (3) a particularized showing of where each claim element is disclosed in the asserted prior art.

35 U.S.C. §§ 103, 322(a)(3), 324(a); *see also KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 418 (2007).

Each of these requirements is independently dispositive, and failure to satisfy any one of them warrants denial as they each independently show that it is not more likely than not that at least one of the claims challenged in the petition is unpatentable. 35 U.S.C. § 324(a).

A. The Petition Does Not Demonstrate a Reasoned Motivation to Combine Prior Art References

Obviousness **requires more than identifying references** that could be combined; a petitioner **must articulate a reasoned motivation to combine those references, supported by a rational underpinning**. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 418 (2007); *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006).

Conclusory assertions—such as invocations of “known techniques”, “minor modifications”, or “predictable results”—are **insufficient** where, as here, the Petition does not explain **how** the specific disclosures would be integrated into a working combination. *See ActiveVideo Networks, Inc. v. Verizon Commc’ns, Inc.*, 694 F.3d 1312, 1327-28 (Fed. Cir. 2012) (rejecting generic expert testimony that “failed to explain how specific references would be combined, which combination(s) of elements in specific references would yield a predictable result, or how any specific combination would operate or read on the asserted claims.”) The Petition repeatedly relies on precisely such conclusory language, asserting that the proposed combinations merely apply “known techniques” (Pet. at 25, 35, 50, 56; Ex1003 ¶¶106, 126, 160, 174), and that any modifications would have required only “minor modifications” (Pet. at 20, 26; Ex1003 ¶88) or yield “predictable results” (Pet. at 19, 25, 35, 50, 56; Ex1003 ¶¶19, 87, 106, 126, 160, 174). Yet these statements are **unaccompanied by any explanation** of **how** the cited references would actually be combined, **what** specific modifications would be made, or **how** the resulting system would operate to meet the claim requirements—precisely the type of unsupported reasoning the Federal Circuit has rejected. *See KSR*, 550 U.S. at 418; *Kahn*, 441 F.3d at 988; *ActiveVideo*, 694 F.3d at 1327-28.

The Board likewise accords little weight to expert testimony like that of Mr. Schmandt (Ex1003), which merely repeats unsupported assertions. *Xerox Corp. v.*

Bytemark, Inc., IPR2022-00624, Paper 9 at 15-17 (PTAB Aug. 24, 2022) (precedential) (holding that the expert testimony should be “entitled to little weight” because the declaration “is conclusory and unsupported” and “adds little to the conclusory assertion for which it is offered to support”).

Petitioner’s argument that it is not offering expert testimony to fill the gap in the prior art is disingenuous. Petitioner’s Response to Patent Owner’s Discretionary Denial Brief, Paper 8 at 13-14. Petitioner asserts that “every limitation is directly mapped” and that Patent Owner has not identified any missing limitation supplied by expert testimony. *Id.* But that is precisely what the Petition does for Examiner identified elements [1(b)] and [1(c)].

During prosecution, the Examiner noted that no cited prior-art reference teaches the pipeline as arranged in the challenged claim. Ex1002 at 132. Petitioner relies on Mr. Schmandt’s testimony to bridge the gap—asserting that *Narayanan*’s alleged generation of “characteristics specific to a natural voice of the second user” can be grafted onto Nguyen’s analysis, while *Narayanan*’s alleged audio-data modification is further combined with *Dirac*’s extracted features and then implemented in Nguyen’s synthesis process, despite the absence of a prior-art teaching or technical roadmap for that stitched-together architecture. Ex1003 ¶¶[121]-[129], [131-134]. However, the references themselves do **not** provide that combined architecture or explain **why** a PHOSITA would have stitched those

disparate teachings together, let alone how they would have resulted in the claimed invention. That is not a teaching in the prior art; it is expert reconstruction. Besides being nonsense, this is exactly the type of expert gap filling that the PTAB cautions against. *Johns Manville Corp. v. Knauf Insulation, Inc.*, IPR2018-00827, Paper 9 at 12-14, 18-20 (PTAB Oct. 16, 2018) (informative) (denying institution and according to the expert declaration little weight where conclusory assertions and expert reliance did not sufficiently explain why the asserted prior-art elements would have been selected and combined to achieve the claimed invention with predictable results).

Here, as discussed in Section V(A), Petitioner does not provide a reasoned explanation for how Nguyen-**Kee-Dirac-Narayanan** would be combined into a coherent system that performs the claimed pipeline, including **[1(b)]** analyzing input data and *generating second-user natural-voice characteristics* and **[1(c)]** synthesizing *modified second-user speech based on those generated characteristics and the extracted accent features*. See Ex1002 at 132; see also **[7(b)]-[7(c)]**, **[14(b)]-[14(c)]**. The Petition therefore fails to establish a legally sufficient motivation to combine.

B. The Petition Does Not Demonstrate a Reasonable Expectation of Success

A petitioner must demonstrate *a reasonable expectation of success for the proposed combination, not merely a motivation to attempt it.* *Intelligent Bio-Sys., Inc. v. Illumina Cambridge Ltd.*, 821 F.3d 1359, 1367-68 (Fed. Cir. 2016); *PAR Pharm., Inc. v. TWI Pharms., Inc.*, 773 F.3d 1186, 1193 (Fed. Cir. 2014). While a combination need not be “the preferred, or the most desirable, combination,” the petitioner must show “there is something in the prior art as a whole to suggest the desirability, and thus the obviousness, of the claimed invention.” *Honeywell Int’l Inc. v. 3G Licensing, S.A.*, 124 F.4th 1345, 1356 (Fed. Cir. 2025) (quotation omitted).

Here, as discussed in Section V(A), Petitioner does not explain **how** the disparate systems of Nguyen-**Kee-Dirac-Narayanan** would **function together** to achieve the claimed integration of pipeline components required by Claim 1, including **[1(b)]** analyzing input data and *generating second-user natural-voice characteristics* and **[1(c)]** synthesizing *modified second-user speech based on those generated characteristics and the extracted accent features*. Nor does the Petition explain **how** these fundamentally different architectures could be (let alone would be) integrated to perform those operations, particularly the Examiner-identified limitation that was found missing from the prior art.

The color coding underscores why that omission is fatal. The Petition does not identify a single prior-art system, or any coherent combination of art, that

performs the claimed sequence. Instead, it assigns the claimed fragments to different references:

- Nguyen for analyzing and synthesizing second input to preserve aspects of the natural voice such that it mimics the first accent,
- **Kee** for the audio interface,
- Dirac for the extracted accent features, and
- *Narayanan* for generating characteristics specific to a natural voice and modifying the obtained second input audio data based on the generated characteristics.

But the claim does not recite these concepts in isolation. It requires an integrated pipeline in which the same speech-derived representations are generated, passed forward, and jointly relied upon across successive processing stages to perform the claimed operations. The Petition, however, treats these steps as modular and interchangeable, sourcing them from different references without explaining how they would be combined into a single, coherent system.

As shown above, independent Claims 7 and 14 recite materially parallel pipeline limitations, and Petitioner relies on the same defective Nguyen-Dirac-*Narayanan* theory for those limitations. Thus, Petitioner's failure to identify or explain this architecture for Claim 1 applies equally to Claims 7 and 14, which recite the same pipeline and therefore rise or fall with the same deficient combination.

The color-coded mapping illustrates not a reasoned combination, but the absence of one: Petitioner has picked disconnected pieces from disparate systems

without showing **how** those pieces would fit together in a working architecture that performs the Examiner-identified limitation, let alone **why** they would be arranged in such a fashion.

In sum, Petitioner never explains **how** the proposed combination would yield the Examiner-identified limitation of [1(b)] analyzing input data and *generating second-user natural-voice characteristics* and [1(c)] synthesizing *modified second-user speech based on those generated characteristics and the extracted accent features*. See Ex1002 at 132.

The Petition therefore fails to establish a reasonable expectation of success.

C. The Petition Does Not Make a Particularized Showing of Where Each Claim Element Is Disclosed in the Asserted Prior Art

The Petition must identify “**with particularity**” where each claim element is disclosed in the prior art and **explain how** the asserted grounds render the claims unpatentable. 35 U.S.C. § 322(a)(3); 37 C.F.R. § 42.204(b)(4).

Here, Petitioner does not identify **any** prior art reference that even purports to disclose the claimed [1(b)] analyzing input data and *generating second-user natural-voice characteristics* and [1(c)] synthesizing *modified second-user speech based on those generated characteristics and the extracted accent features*, or a system that performs the claimed pipeline as a whole. Instead, it attempts to cobble together that limitation from separate and disparate references— Nguyen-*Kee-Dirac-Narayanan*—without demonstrating that the combined system forms a coherent pipeline or performs the claimed sequence of operations, including the allowance-dependent limitation.

Apple Inc. v. Ginko LLC underscores that, where, as here, a petition fails to map claim elements with particularity, denial at institution is warranted. IPR2025-01388, Paper 14 at 3 (PTAB March 2, 2026). There, the Director emphasized that a petitioner must, “from the onset,” provide an understandable, element-by-element explanation grounded in the prior art. *Id.* Petitioner’s showing here falls short for the same reasons: it neither identifies a prior-art disclosure of the claimed pipeline, nor

explains how the proposed combination performs **that requirement within the claimed pipeline.**

D. Institution Should Be Denied as an Inefficient Use of Office Resources

Finally, denial is also warranted where, as here, institution would not serve the efficient administration of the Office. The Director has confirmed that the Board may deny institution where only a subset of claims or grounds has potential merit and proceeding to trial would waste resources. *CyberSecure IPS, LLC v. Network Integrity Systems, Inc.*, IPR2025-01441, Paper 18 at 2-3 (PTAB May 26, 2026) (citing *Chevron Oronite Co. v. Infineum USA*, IPR2018-00923, Paper 9 at 10-11 (PTAB November 7, 2018) (informative)).

Petitioner's reliance on the *Sotera*-style stipulation (Petitioner's Response to Patent Owner's Discretionary Denial Brief, Paper 8 at 6-7) does not meaningfully mitigate this inefficiency. Although Petitioner purports that it will not pursue in the district court litigation "the specific grounds raised in PGR2026-00033" (Ex1037 at 2), the submitted invalidity contentions confirm that Petitioner continues to assert broad combinations of references (including the art presented in this proceeding, along with other art), system-based theories, and expert-driven PHOSITA reasoning that fall outside the scope of any estoppel. Ex2033 at 4-13. For example, Petitioner expressly reserves the ability to mix and match prior art across references and systems, expanding the scope of its invalidity case beyond the grounds asserted in

the Petition and undermining the intended narrowing effect of Sotera-style stipulations:

The references in the below charts may be combined with one or more references in the same chart, another chart, or another piece of prior art identified in these contentions, for at least the same exemplary reasons provided in the motivation to combine sections of the charts and Krisp’s IPR and PGR Petitions, which are incorporated by reference. IPR2026-00275 (’457 Patent); PGR2026-00032 (’561 Patent); IPR2026-00273 (’496 Patent); IPR2026-00272 (’550 Patent); IPR2026-00274 (’745 Patent); PGR2026-00033 (’756 Patent).

Ex2033 at 10 (emphasis added).

Thus, the same underlying factual disputes, expert testimony, and technical theories will proceed in parallel regardless of institution. The stipulation alters only the labeling of Petitioner’s theories—not their substance—leaving intact duplicative proceedings that do not simplify the issues before the parties or the Board. Moreover, Petitioner continues to assert invalidity and reserved broad rights to amend and supplement its invalidity positions. Ex2033 at 2-3. These reservations ensure that the district court litigation will proceed in full irrespective of any stipulation, further confirming that institution would not reduce the scope of the dispute and would result in inefficient use of the Board’s resources.

Here, all asserted grounds depend on the same defective Nguyen-[Dirac-Narayanan](#) and/or Nguyen-[Kee-Dirac-Narayanan](#) combinations, and the additional

references in Grounds 2-6 do not cure the deficiencies in the independent claims. Instituting trial would therefore require the Board to expend resources on multiple grounds that fail to address the dispositive issue.

V. THE PETITION FAILS TO SHOW THAT ANY CHALLENGED CLAIM IS MORE LIKELY THAN NOT UNPATENTABLE

Petitioner asserts six obviousness grounds:

Ground	Claims Challenged	References Relied Upon	Petition Citation
Ground 1	1-3, 6	Nguyen, Kee , <u>Dirac</u> , <i>Narayanan</i>	Pet. at 12
Ground 2	4	Nguyen, Kee , <u>Dirac</u> , <i>Narayanan</i> , Kamruzzaman	Pet. at 46
Ground 3	5	Nguyen, <u>Dirac</u> , <i>Narayanan</i> , Wang	Pet. at 51
Ground 4	7-11, 13	Nguyen, <u>Dirac</u> , <i>Narayanan</i> , Kamruzzaman	Pet. at 57
Ground 5	14-17, 19-20	Nguyen, <u>Dirac</u> , <i>Narayanan</i> , Wang	Pet. at 60
Ground 6	12, 18	Nguyen, <u>Dirac</u> , <i>Narayanan</i> , Kamruzzaman , Wang	Pet. at 63

This Preliminary Response focuses on Ground 1 because Ground 1 supplies the base combination that every other asserted ground, and every claim, requires.²

² Grounds 2-6 rely on the same defective Nguyen-Dirac-*Narayanan* core theory for the independent-claim pipeline limitations, with **Kee**, **Kamruzzaman**, or **Wang** added or omitted depending on the particular claim form and dependent limitation.

Ground 1 asserts that Claims 1-3 and 6 are obvious over Nguyen, [Kee](#), [Dirac](#), and [Narayanan](#). Pet. at 12. Grounds 2-6 merely add [Kamruzzaman](#) or [Wang](#) or remove [Kee](#), for dependent-claim limitations, while retaining the same Nguyen-[Kee](#)-[Dirac](#)-[Narayanan](#) or Nguyen-[Dirac](#)-[Narayanan](#) combination for the independent-claim limitations. Pet. at 46, 51, 57, 60, 63. Thus, if Ground 1 fails to show the claimed alignment in independent Claims 1, 7, and 14, Grounds 2-6 fail as well.

A. Ground 1 Fails on the Merits

The claim elements specifically identified by the Patent Office are embodied in claim elements [1(b)]-[1(c)], [7(b)]-[7(c)], and [14(b)]-[14(c)]. Claim [1(b)] requires “analyz[ing] obtained second input audio data associated with second input speech associated with a second accent of a second user to generate characteristics specific to a natural voice of the second user, wherein the generated characteristics correspond to vocal traits that are distinct to the second user and comprise one or more of a voice quality or one or more phonetic patterns, prosodic features, articulation styles, or intonation patterns.” Ex1001 Cl. 1; *see also* Ex1001 Cls. 7, 14.³ Petitioner’s Ground 1 combination does not satisfy these requirements. Claim

³ For Claim 7, Ground 4 adopts the same analysis as Ground 1, except it relies upon Ground 2, Claim 4, for the additional limitation of “unique fingerprint of the voice of the second user”. Pet. at 57-59. Similarly, for Claim 14, Ground 5 adopts the same

[1(c)] requires “synthesiz[ing] a modified version of the second input speech by modifying the obtained second input audio data based on the generated characteristics and the extracted accent features, wherein the modified version of the second input speech preserves aspects of the natural voice of the second user and mimics the first accent;” Ex1001 Cl. 1; *see also* Ex1001 Cls. 7, 14. Petitioner’s Ground 1 combination does not satisfy these requirements.

Petitioner’s theory is not that any reference teaches the claimed speech processing path. Instead, Petitioner divides the limitation across **four** references. Petitioner relies on Nguyen for the alleged analyzing and synthesizing second input to preserve aspects of the natural voice such that it mimics the first accent. Pet. at 27-30, 37-39. Petitioner then turns to *Narayanan* for generating characteristics specific to a natural voice and modifying the obtained second input audio data based on the generated characteristics. Pet. at 30-33, 39-40. Additionally, the Petition relies on *Dirac* for the extracting accent features and *Kee* for the accompanying audio interface that these components interface with. Pet. at 12, 16-17, 22-24. Each reference addresses different data, objectives, and architectures and the Petition fails

analysis as Ground 1, except it relies upon Ground 3, Claim 5, for the additional limitation of “speaker-specific voice characteristics”. Pet. at 60-62.

to demonstrate how the proposed combination would overcome these differences and result in the claimed invention.

The division of labor proposed by the Petition is not a teaching in the prior art; it is a purported reconstruction from the claim language. Obviousness requires more than assigning pieces of a claim to different references. *See KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 418 (2007); *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006). Petitioner had to show a coherent combined system **that combines first-user accent features with second-user natural-voice characteristics to generate output audio that preserves aspects of the second user's voice while mimicking the first user's accent.** Ex1001 Cl. 1. It did not.

Accordingly, the challenges to independent Claims 1, 7, and 14 fail on the merits.

1. The Ground 1 Combination is Architecturally Incoherent

Here, the Petition fails because it never shows the claimed integrated accent-mimicking pipeline. The reason the Petition cannot make that showing is that the alleged Nguyen-**Kee-Dirac-Narayanan** combination is inoperable. Petitioner does not identify a coherent prior-art system that performs the claimed sequence of operations, which the Patent Office specifically called out as inventive. Instead, it attempts to assemble Nguyen's high-level accent conversion, **Kee's** audio interface, **Dirac's** accent-characteristic analysis and accent-translation models, and

Narayanan's same-speaker word-correction voice profile. But no amount of additional explanation can reconcile the defects in that proposed combination.

For the generated-characteristics step **[1(b)]**, the proposed Nguyen-*Narayanan* combination fails to supply the claimed architecture. *Accord* Ex1001, Cls. **[7(b)]**, **[14(b)]**. Claim **[1(b)]** requires analyzing second-user audio to generate characteristics specific to the second user's natural voice. Petitioner relies on Nguyen for retaining speaker identity characteristics of speech during accent conversion and on *Narayanan* for generating a voice profile. Pet. at 27-37. But Nguyen's disclosure is high-level: it states that its AC model converts speech from a source accent to a target accent "while retaining the identity characteristics of the original speaker," without explaining how those characteristics are generated or supplied to the model. Ex1004 ¶¶[0013]-[0014], [0072]-[0073]. *Narayanan*, by contrast, is not an accent-conversion system. It is a same-speaker audio-correction system that identifies incorrect words, generates corrected replacement words based on the speaker's voice profile, and inserts those corrections into the same speaker's content. Ex1006 ¶¶[0004]-[0005], [0018], [0075], [0083]-[0084]. Petitioner never explains how *Narayanan*'s segmented word-correction voice profile becomes an operative input to Nguyen's stream-based accent-conversion system. At most, the Petition identifies that *Narayanan* can create a voice profile; it does not show how

that profile is integrated into Nguyen’s AC model so that it performs the claimed generation-and-synthesis pipeline.

For the synthesis step [1(c)], the proposed combination fails twice over. First, the Petition does not show synthesis based on the first user’s extracted accent features. *Accord* Ex1001 Cls. [7(c)], [14(c)]. Claim [1(c)] requires synthesizing a modified version of the second input speech “based on the generated characteristics and the extracted accent features.” Ex1001 Cl. [1(c)]. Petitioner’s theory is that Dirac’s audio characteristics are used to detect accents and select Nguyen’s AC model, and that “because the AC model is selected based on the extracted accent features, the output of the AC model is based on the extracted accent features.” Pet. at 39 (citing Ex1003 ¶ 134 (Schmandt Decl.)). The Petition makes this assertion not based upon any disclosure in Nguyen or Dirac, but instead on Mr. Schmandt’s unsupported conjecture. If anything, this is an admission that the prior art does not disclose limitation [1(c)], and that instead the only way to reconstruct it from the prior art is to use the patent as a guide, which is impermissible hindsight.

Second, the Petition does not show synthesis based on the second user’s generated voice characteristics. The Petition maps the generated characteristics to Narayanan’s voice profile, but Nguyen’s AC model provides no disclosed input path through which that profile could inform synthesis. Nguyen’s AC model is “based on an encoder-decoder paradigm with an attention mechanism,” in which the encoder

receives a “mel-filterbank with 80 features” and the decoder generates converted speech from the encoded audio. Ex1004 ¶¶[0077]-[0078]. Nguyen does not disclose a separate runtime input for a speaker profile, voice-characteristic vector, embedding, latent representation, or other conditioning signal.

Kee does not cure any of those defects. Petitioner adds **Kee** only for the audio-interface hardware in the preamble: an audio interface connected to a microphone and speakers, including digital-to-analog and analog-to-digital converters. Pet. at 16-20; Ex1012 ¶[0027].⁴ **Kee** says nothing about applying trained ML models to extract accent features, generating second-user natural-voice characteristics, or synthesizing modified second-user speech based on both first-user accent features and second-user voice characteristics. It supplies generic hardware, not the claimed speech-processing pipeline. The Petition’s reliance on **Kee** therefore does not make the Nguyen-Dirac-*Narayanan* architecture coherent.

⁴ The Petition relies on **Kee** for the audio interface limitation of claim 1. Pet. at 16-17. However, no other claim recites an audio interface limitation, and the Petition correspondingly does not include **Kee** for the remaining independent claims or their asserted grounds. For the reasons discussed herein, the problems with the combination of Nguyen, Dirac and Narayanan are fatal to the petition, whether or not they are combined with **Kee** for the other Grounds.

The proposed combination also fails because the references perform different operations on different data for different purposes. Nguyen is an end-to-end virtual-conference accent-conversion system that broadly converts source-accent audio to target-accent audio while retaining speaker identity. Ex1004 ¶¶[0013], [0072]. [Dirac](#) identifies or translates accents using accent sample sets and translation models. Ex1005, Abstract, 2:40-45, 6:4-7:13, Fig. 3. [Narayanan](#) performs localized same-speaker audio correction by replacing incorrect words with corrected words generated from the same speaker’s voice profile. Ex1006 ¶¶[0004]-[0005], [0018], [0083]-[0084]. [Kee](#) provides conventional audio-interface hardware. Ex1012, ¶[0027]. The Petition never explains **how** those components would be integrated into a single system in which first-user accent features and second-user natural-voice characteristics are both operative synthesis inputs.

Nor does the Petition establish a reasonable expectation of success. Its repeated assertions that a PHOSITA would make only “minor modifications” do not explain **what** the combined system actually is, which model performs synthesis, **how** [Dirac](#)’s extracted accent features are supplied to that model, or **how** [Narayanan](#)’s voice profile conditions Nguyen’s AC model. Pet. 19-20, 24-27, 33-37, 49-51, 55-57. In particular, the Petition does not identify any data path by which [Narayanan](#)’s generated voice characteristics become usable by Nguyen’s audio-to-audio model, nor any input through which [Dirac](#)’s first-user extracted accent features drive

Nguyen's modification of second-user speech. The resulting theory is not an integrated speech-processing architecture; it is a collection of disconnected disclosures selected after the fact to mirror the claim language. **This failing is fatal to all of the grounds advanced in the Petition.**

Accordingly, the alleged Nguyen-**Kee-Dirac-Narayanan** combination does not teach or render obvious Claim 1's integrated pipeline. Nguyen supplies a high-level accent-conversion framework, **Kee** supplies generic audio hardware, **Dirac** supplies accent-characteristic analysis and accent translation concepts, and **Narayanan** supplies a same-speaker voice profile for word correction. The Petition does not provide an operable path for those components to work together to extract first-user accent features with a trained ML model, generate second-user natural-voice characteristics, and synthesize modified second-user speech based on both. Because Petitioner fails to show a coherent architecture, a reasoned motivation to combine, or a reasonable expectation of success, Ground 1 fails.

Each of these incompatibilities independently renders the proposed combination non-operational for the claimed purpose. The result is not an integrated system that extracts, generates, and synthesizes as Claim 1 requires, but a set of disconnected components that cannot be assembled into the claimed invention without rebuilding the system using the '756 Patent as a roadmap. Those failures

carry through to Claims 2, 3, and 6, which depend from Claim 1, and to the later grounds that rely on the same defective Nguyen-[Dirac](#)-[Narayanan](#) core.

2. The Nguyen-Dirac Combination Does Not Teach Applying a Trained Machine Learning Model to Extract Accent Features

Claim [1(a)] requires the system to “*apply one or more trained machine learning models to first input audio data ... to extract accent features of first input speech associated with a first accent of a first user.*” Ex1001, Cl. [1(a)]; *see also* [7(a)], [14(a)]. Petitioner fails to identify any disclosure in the provided references that satisfies the requirement. Its theory instead depends on combining distinct aspects of [Dirac](#)’s system:

- (1) a neural-network-assisted generation/refinement of accent translation models;
- (2) an audio-characteristic analysis and correlation process; and
- (3) the resulting accent translation model.

See Pet. at 22-27.

None of those disclosures teaches the claimed operation. [Dirac](#) discloses audio-characteristic analysis but does **not** identify a trained ML model as performing that analysis on first input audio. [Dirac](#)’s neural-network disclosure is directed to generating, updating, or refining accent translation models, **not** to extracting features from first input audio; and [Dirac](#)’s accent translation model is used for accent conversion, **not** feature extraction. 1:66-2:54, 10:2-12.

The neural network the Petition invokes in [Dirac](#) is a model-*building* tool, **not** an actual model applied to input audio. [Dirac](#) explains that “the accent translation system may continually and/or repeatedly receive new audio samples and use these new audio samples to update and refine the accent translation models,” and that “the generating, updating, and/or refining of the accent translation models may be performed at least in part by one or more artificial neural networks ... that may efficiently parse and analyze large amounts of input data.” Ex1005, 2:45-54. Thus, in [Dirac](#), the neural network receives a corpus of audio samples and outputs or refines an accent translation model. It is **not** applied to the first user’s input audio to extract accent features.

Further, [Dirac](#) utilizes non-machine-learning modules to evaluate characteristics of the first party’s input speech. [Dirac](#) describes analyzing the first party’s audio communications to obtain audio characteristics—such as pitch, tone, stress, and melody—using processing techniques including FFT and MFCC, and then comparing those characteristics to accent sample sets to determine a correlated accent. Ex1005, 8:59-9:2. Although [Dirac](#) separately discloses neural networks for generating, updating, or refining accent translation models, it does **not** attribute the FFT/MFCC-based audio-characteristic analysis or comparison step to any trained neural network.

Nor does [Dirac](#)'s trained accent translation model satisfy the limitation. That model is used for conversion, **not** extraction. [Dirac](#) describes the accent translation model as including "instructions for translating audio such that audio characteristics associated with the first accent sample set are adjusted to more closely resemble those of the second accent." Ex1005, 2:40-44. Adjusting audio toward a target accent is an accent-conversion function; it is **not** the claimed extraction of accent features from first input audio data.

Thus, [Dirac](#) does not satisfy Claim **[1(a)]**. The Petition constructs the limitation only by taking the "feature extraction" aspect from [Dirac](#)'s signal-processing analysis and the "machine-learning" aspect from [Dirac](#)'s model-generation process, then attributing both to a single operation that [Dirac](#) never discloses. Pet. at 20-22. That purported reconstruction is improper and does not meet the claim requirement that a trained machine-learning model be applied to first input audio data to extract accent features.

3. The Proposed Nguyen-Kee-Dirac-Narayanan Combination Does Not Teach Synthesizing the Second User's Speech Based on the First User's Extracted Accent Features

The Petition does not identify any reference, or put forth any adequately supported combination, that synthesizes the second user's modified speech **based on the first user's extracted accent features**. Claim **[1(c)]** requires "synthesiz[ing] a modified version of the second input speech by modifying the obtained second

input audio data based on the generated characteristics and the extracted accent features.” Ex1001, Cl. [1(c)]; *see also* [7(c)], [14(c)]. The “extracted accent features” are the first user’s features identified with respect to Claim [1(a)], and the claim requires those features to inform the modification of the second user’s speech. Nguyen and [Dirac](#), on which the Petition relies for this proposition, do not disclose—either alone or in combination—a synthesizer that receives or uses a different user’s extracted accent features as a basis for modifying the second user’s speech.

The audio characteristics obtained from the first party’s speech in [Dirac](#) are used to determine a correlated accent: they “may then be compared to audio characteristics of various accent sample sets ... to determine an accent set and an accent to which the first party’s accent substantially correlates.” Ex1005, 8:64-9:2. After that accent is determined, translation is performed using an accent translation model directed to accent sample sets—**not** to the first party’s specific extracted features. Ex1005, 2:40-44, 9:39-65. [Dirac](#) states that “audio characteristics associated with the first accent sample set are adjusted to more closely resemble those of the second accent audio sample set” (Ex1005, 2:40-44), and further explains that the model may adjust first-accent pitch, tone, stress, and melody to resemble corresponding second-accent characteristics. Ex1005, 9:39-65. Thus, once first-party speech is mapped to an accent class, [Dirac](#) does **not** disclose any further role for that user’s individual extracted characteristics in the downstream conversion.

The conversion is governed by the selected accent sample-set model, not by the first user's extracted accent features.

Moreover, substituting the first user's Dirac-extracted accent features for Dirac's accent-sample-set translation model is **not** a combination a PHOSITA at the time of invention would have made, because it works against the very design of Dirac's system. Dirac's accent translation model is generated by comparing accent sample sets, **not** by using one speaker's individual feature values; those sample sets are built from multiple speakers and may even be combined into a "neutral accent." Ex1005, 2:7-64. Replacing Dirac's translation model with one speaker's idiosyncratic features would undermine Dirac's speaker-independent accent translation and risk conflating accent with the first speaker's voice identity. Further, Dirac permits accent determination by non-audio proxies such as area code or manual selection, confirming that individualized extracted features are **not** required synthesis inputs. Ex1005, 8:33-52. Using one speaker's extracted pitch, tone, stress, or melody values would undermine Dirac's class-level, speaker-independent translation scheme and risk transferring the first speaker's idiosyncratic vocal traits rather than the accent-class characteristics Dirac's model is designed to capture.

While the Petition reasons that "because the AC model is selected based on the extracted accent features, the output of the AC model is based on the extracted accent features" (Pet. at 39 (citing Ex1003 ¶134)), using features as labeling input

to select a model is **not** equivalent to modifying the second user’s speech based upon those features. Claim [1(c)] requires synthesis “based on ... the extracted accent features,” such that the extracted features must inform the modification itself; otherwise, the limitation collapses into mere accent detection or model selection. The Petition’s combination removes “the extracted accent features” from the synthesis step, and the expert declaration (Ex1003 ¶134) merely repeats the conclusion without identifying **where** the references teach feature-based synthesis. Conclusory expert testimony cannot supply a missing disclosure. Pet. at 37-40.

Narayanan does not cure the deficiency because it uses speaker characteristics for a different purpose. *Narayanan* generates replacement audio to correct a speaker’s own words, using that same speaker’s voice profile so the correction sounds like the speaker actually said it. Ex1006 ¶¶[0004]-[0005], [0083]-[0084]. It does **not** take accent features from one user and use them to modify another user’s speech. The same is true in *Narayanan*’s multiple-speaker examples: the system separates the speakers and corrects each speaker’s audio using that speaker’s own profile. Ex1006 ¶[0051]. *Narayanan*’s brief reference to accent is also limited to correction, **not** accent transfer. It describes replacing a “non-standard accent” with a “normalized accent” version, again as part of correcting the same speaker’s speech—**not** extracting one user’s accent features and applying them to another user’s speech. Ex1006 ¶[0055].

Thus, the Petition’s combination does not disclose synthesis based on the first user’s extracted accent features. Nor does the Petition explain **why** or **how** a PHOSITA at the time of invention would modify Nguyen’s label-selected conversion, Dirac’s classify-then-select process, and *Narayanan*’s same-speaker correction to arrive at the claimed feature-driven synthesis.

4. Nguyen’s Architecture Provides No Path for *Narayanan*’s Voice Profile to Inform Synthesis

The combination, as articulated by Petitioner, is non-operational for the claimed purpose. Claim [1(c)] requires the second user’s generated voice characteristics to inform synthesis. The “generated characteristics” are the second user’s speaker-specific voice characteristics used to preserve that user’s voice identity, distinct from the “extracted accent features,” which are accent features of the first user’s accent used to supply the accent to be mimicked.

As disclosed, Nguyen’s runtime input is audio; it identifies **no** separate branch or conditioning input for a speaker profile, voice-characteristic vector, embedding, latent representation, **nor** other speaker-characteristics input.

The Petition articulates **no** place in Nguyen’s pipeline where *Narayanan*’s voice profile would attach, and its assertion that “the second audio stream would have been based on the generated voice profile” (Pet. at 39), states a result **without** identifying the mechanism. Claim [1(c)] is also unmet.

The Petition’s own positions confirm that speaker characteristics **cannot** simply be handed to a machine-learning synthesizer and assumed to affect the output. In later grounds, the Petition states that “in order for the identity characteristics to be usable by a machine learning model, the identity characteristics must be first encoded in a usable medium,” and adds Wang to “encode[] the identity characteristics into a speaker embedding that is usable by the processing system.” Pet. at 55-56.

Ground 1, however, includes **no** such encoding step and **no** conditioning input in Nguyen capable of receiving Narayanan’s voice profile, yet asserts that Nguyen’s model synthesizes audio data “based on” that raw profile. Pet. at 39-40. This is not sufficient.

Nor would adding Wang’s encoder alone cure the defect: Wang’s encoder works with Wang’s own generator, which is designed to receive the resulting speaker embedding, whereas Nguyen’s AC model is an audio-to-audio encoder-decoder system with **no** disclosed input for such an embedding. Storing an embedding “in the voice profile” (Pet. at 55-56), does **not** give Nguyen’s model any mechanism to use it.

Thus, whether the alleged generated characteristics are Narayanan’s raw voice profile or Wang’s encoded speaker embedding, the Petition identifies **no** synthesizer in the asserted combination capable of receiving those characteristics

and using them to synthesize modified speech “based on the generated characteristics” as Claim [1(c)] requires. This is also fatal to the Petition.

5. Ground 1 Dependent-Claim Challenges Do Not Cure the Independent-Claim Defects

Ground 1 also challenges dependent Claims 2, 3, and 6. The Petition relies on the same Nguyen-**Kee-Dirac-Narayanan** combination for those claims. The Petition’s dependent-claim theories do not cure the deficiencies outlined above in the independent claims, and rather perpetuate and expand upon the incoherent patchwork combination of Claim 1. Because Petitioner fails to establish the limitations of Claim 1 and fails to cure these deficiencies in the independent claims, its challenges to all dependent claims necessarily fail.

Accordingly, Ground 1 fails for all claims challenged in that ground.

B. Grounds 2-6 Fail on the Merits

1. Ground 2 Fails Because Kamruzzaman’s MFCC Speaker-Identification Features Do Not Teach a Synthesis-Driving Voice Fingerprint

Ground 2 adds **Kamruzzaman** to supply Claim 4’s MFCC-fingerprint limitation: the Petition relies on **Kamruzzaman**’s MFCC analysis and resulting MFCC-derived support vector machine (“SVM”) speaker representation as the alleged “unique fingerprint” of Nguyen’s second participant’s voice, and then treats that fingerprint as incorporated into **Narayanan**’s voice profile—the alleged “generated characteristics.” Pet. at 46-51.

Ground 2 does not offer any fix for the failures of Ground 1 discussed above.

In any event, because Claim 4 depends on Claim 1, that fingerprint must be part of the generated characteristics used to “synthesize a modified version of the second input speech ... *based on the generated characteristics*” while “preserv[ing] aspects of the natural voice of the second user.” Ex1001 Cl. 1 (emphasis added). Thus, the claimed MFCC-derived fingerprint must **do more than merely identify** the speaker; it must be part of the voice characteristics that inform synthesis and preserve the second user’s natural voice. **Kamruzzaman does not teach that use.**

Kamruzzaman is directed to speaker identification for “authentication and verification” in security applications. Ex1011 at 1. It uses MFCC-based voice features **to tell speakers apart, not to reproduce** a speaker’s voice. Those features are fed into a classifier that identifies which speaker produced a given utterance. Ex1011 at 2. Although **Kamruzzaman** states that the features can “represent and recognize the voice characteristic of the speaker” (Ex1011 at 2), recognizing a speaker is **not the same** as synthesizing that speaker’s voice. **Kamruzzaman** never describes using MFCC features to synthesize, generate, or reconstruct speech. A voice feature used to identify who is speaking does **not**, without more, teach a fingerprint used to drive voice-preserving synthesis.

That mismatch compounds the structural deficiency already identified for Claim 1. Even assuming **Kamruzzaman**'s MFCC analysis produces a "fingerprint," the Petition **does not identify any synthesizer** in the asserted combination that can use that fingerprint to preserve the second user's voice. Nguyen's AC model is an encoder-decoder system that receives input audio features and outputs converted audio; the cited architecture discloses encoder inputs, attention-based decoding, and mel-filterbank/spectrogram prediction, but **does not disclose a separate input** through which an extracted fingerprint could condition synthesis. Ex1004 ¶¶[0077]-[0079], [0083], [0112]-[0115]. Incorporating **Kamruzzaman**'s MFCC features into *Narayanan*'s voice profile, as the Petition proposes (Pet. at 46-51) does **not** solve that problem because the combination still lacks a disclosed mechanism by which those features reach and control Nguyen's synthesis process.

The same mismatch undermines the Petition's motivation theory. The Petition makes no effort to demonstrate that a PHOSITA at the time of the invention seeking to preserve a speaker's voice during accent conversion would be led to a speaker-authentication reference whose disclosed purpose is to distinguish and verify speakers for security. Pet. at 49-51. **Kamruzzaman** addresses whether an utterance belongs to a particular speaker; it **does not teach extracting a representation from which that speaker's voice can be re-synthesized or preserved during accent conversion**.

The Petition’s rationale—that *Narayanan* lacks implementation details for determining and representing voice traits (Pet. at 49-51)—identifies a combination gap, but does **not** adequately explain **why** a speaker-recognition technique would supply the claimed synthesis-driving fingerprint. Petitioner therefore has **not** shown that the proposed combination teaches extracting a unique fingerprint that forms part of generated characteristics used to synthesize modified speech while preserving the second user’s voice.

2. Ground 3 Fails Because Wang’s Encoding Step Does Not Provide a Synthesizer That Uses the Encoded Speaker Characteristics

Ground 3 adds *Wang* to supply Claim 5’s speaker-identity-encoding limitation: the Petition relies on *Wang*’s speaker encoder to encode *Narayanan*’s speaker-specific voice characteristics into speaker embedding data, and then treats that embedding as incorporated into *Narayanan*’s voice profile—the alleged “generated characteristics.” Pet. at 51-57.

Ground 3 does not offer any fix for the failures of Ground 1 discussed above.

In any event, Claim 5 requires the system to “apply a speaker identity encoding technique to encode speaker-specific voice characteristics, wherein the generated characteristics comprise the speaker-specific voice characteristics.” Ex1001 Cl. 5. Because Claim 5 depends on Claim 1, those encoded speaker-specific

characteristics must also be used to “synthesize” the modified speech “based on the generated characteristics.” *Id.* at Cl. 1. Thus, it is not sufficient to encode speaker characteristics in isolation; the encoded characteristics must be available to, and used by, the synthesizer. The Petition disconnects the encoding requirement from the synthesis requirement. Ground 3 also illustrates the Petition’s inconsistent treatment of its own base combination. The Petition relies on the Ground 1 Nguyen-**Kee-Dirac-Narayanan** combination for independent Claim 1, from which Claim 5 depends, but its Ground 3 heading omits **Kee** when adding **Wang** for Claim 5. Although **Kee** does not cure any substantive defect because it is used only for generic audio-interface hardware, the omission underscores that **Ground 3 merely layers Wang onto an already incomplete and inconsistently articulated combination.**

As explained above with respect to Claim 1, Nguyen’s AC model does not disclose an input capable of receiving an encoded speaker representation. Nguyen’s AC model is an encoder-decoder accent-conversion system that operates on audio inputs and outputs mel-filterbank features for vocoding, preserving speaker identity through paired training data; Nguyen **does not disclose a speaker-embedding input or other conditioning pathway** by which an encoded speaker representation is supplied to the AC model. Ex1004 ¶¶[0076]-[0083].

Adding **Wang** supplies an encoder, **but not a synthesizer in Nguyen that can use the encoder’s output.** **Wang’s** speaker embedding is designed for **Wang’s** own

expressive speech generator, which is configured to receive that embedding and use it to generate output audio. Ex1007 ¶¶[0047]-[0049]. Nguyen’s AC model, by contrast, is **not disclosed as accepting or using such an embedding**. *See* Ex1004 ¶¶[0077]-[0079], [0083]. The Petition therefore produces an encoded embedding and then merely places it “in the voice profile of the second user.” Pet. at 55-57. But storing an embedding in a profile is **not the same** as supplying it to a synthesizer that can act on it.

Accordingly, the Petition’s reading of Claim 5 is completely isolated: the embedding is generated and stored, but nothing in the asserted combination consumes it during synthesis. Encoding changes the format of the speaker characteristics; it **does not give Nguyen’s model the ability to use them**. The inherited Claim 1 requirement remains unmet because the Petition still **does not identify any synthesis process** that operates “based on” the encoded speaker-specific characteristics.

That failure also defeats the Petition’s stated rationale for adding Wang. The Petition adds Wang on the theory that “in order for the identity characteristics to be usable by a machine learning model, the identity characteristics must be first encoded in a usable medium.” Pet. at 55-56. But usability depends on the model that must use the encoded result. Here, the relevant model is Nguyen’s AC model, and the Petition **does not show that Nguyen’s model can receive or use a speaker**

embedding (let alone would). Wang may explain **how** to encode identity characteristics, but the Petition does **not** supply the corresponding mechanism that makes Nguyen’s synthesizer consume those encoded characteristics.

The result is an architecturally incomplete combination: it produces an embedding **without** a disclosed path for that embedding **to affect synthesis**. This is fatal to Petitioner’s analysis, and therefore has not shown that the Wang combination satisfies Claim 5’s encoding requirement together with the inherited Claim 1 requirement that synthesis be performed “based on the generated characteristics.”

3. Ground 4 Fails Because Nguyen’s Real-Time Media Transport Does Not Teach Real-Time Accent Conversion

Ground 4 relies on Nguyen to satisfy Claim 7’s “real-time” requirement. Claim 7 is directed to “[a] method for real-time accent mimicking,” a requirement that appears in no other independent claim. Ex1001 Cl. 7. Neither of these teachings fix the fundamental flaws in Ground 1. Thus, **Ground 4 does not offer any fix for the failures of Ground 1 discussed above.**

In any event, for Ground 4, Petitioner must show that the asserted combination performs accent conversion in real time. It has not done so. The only “real-time” disclosure the Petition identifies concerns Nguyen’s conferencing media-transport infrastructure, **not the accent-conversion processing itself.**

The Petition’s support is that conference communications occur through a “real-time media server,” from which it concludes that a PHOSITA “would have

understood that Nguyen’s accent conversion method occurs in real-time.” Pet. at 57-58 (citing Ex1004 ¶¶[0052]-[0054]; Ex1003 ¶178). But the Nguyen citations concern conference administration and infrastructure, **not** the operation of the accent-conversion model. They describe functions such as admitting participants, managing hosts, and creating breakout rooms, while “real-time media servers” are part of the infrastructure that routes multimedia streams.

Real-time transport of media streams does not establish real-time accent conversion of those streams. Nguyen never states that its accent-conversion model processes live speech in real time. Indeed, the only conversion-related discussion using “real-time” concerns sample collection: native-accent recordings may be stored in memory, and “in some examples” speech samples may be “received in real-time from the speakers” for processing by a voice-conversion model. Ex1004 ¶¶[0099]-[0100]. Receiving speech samples in real time is not the same as converting live speech in real time. The Petition therefore infers real-time conversion from real-time transport, which does not follow from the cited disclosure.

Nguyen’s disclosed model architecture reinforces the gap. Nguyen describes its accent-conversion model as “based on an encoder-decoder paradigm with an attention mechanism,” using a “pyramid bi-LSTM” encoder. Ex1004 ¶¶[0077]-[0078]. The Petition **does not identify** any disclosure that this model operates in real time. **Nor does it explain** how Nguyen’s bidirectional, attention-based model would

satisfy the claimed real-time requirement. At most, the Petition shows that Nguyen’s conferencing platform transmits media in real time; it does **not** show that Nguyen’s accent-conversion processing is performed in real time. Petitioner therefore has not shown that the asserted combination teaches the “real-time accent mimicking” required by Claim 7 or its dependent claims.

4. Ground 5 Fails Because the Petition Merely Repackages the Same Defective Claim 1 and Claim 5 Theories for the Computer-Readable-Medium Claims

Ground 5 combines Nguyen, [Dirac](#), [Narayanan](#), and [Wang](#). Pet. at 60-63. Independent Claim 14 parallels Claim 1, adding the requirement that the system be executed on a “non-transitory computer-readable medium.”

Ground 5 does not offer any fix for the failures of Ground 1 discussed above.

In any event, the Petition’s Ground 5 analysis largely cross-references its Grounds 1 and 3 arguments. *See* Pet. at 60-63. Those cross-references carry forward the same defects. Petitioner still offers no reasoned explanation for how a PHOSITA would integrate Nguyen’s stream-based accent conversion, [Narayanan](#)’s segmented word-correction voice profile, [Dirac](#)’s accent-feature analysis, [Kamruzzaman](#)’s MFCC speaker-identification approach, and [Wang](#)’s speech-translation embeddings into a functioning real-time accent-mimicking system.

5. Ground 6 Fails Because It Merely Aggregates the Same Defective Theories Without a Coherent Accent-Mimicking Architecture

Ground 6 combines Nguyen, [Dirac](#), [Narayanan](#), [Kamruzzaman](#), and [Wang](#) for Claims 12 and 18. Pet. at 63-64. Claim 12 depends from Claim 7 and adds speaker-identity encoding; Claim 18 depends from Claim 14 and adds MFCC analysis to extract a unique fingerprint.

Ground 6 does not offer any fix for the failures of Ground 1 discussed above.

In any event, Ground 6 merely aggregates the defective rationales from Grounds 1 through 5. It does **not** explain **why** a PHOSITA would combine Nguyen's stream-based accent conversion, [Narayanan](#)'s segmented word-correction voice profile, [Dirac](#)'s accent-feature analysis, [Kamruzzaman](#)'s MFCC speaker-identification approach, and [Wang](#)'s speech-translation embeddings into a coherent real-time accent-mimicking system. Petitioner's reliance on additional references only underscores the hindsight-driven nature of the reconstruction.

6. Grounds 2-6 Fail Because They Rely on Ground 1

Grounds 2-6 fail at least because they depend on Ground 1's defective base combination of Nguyen-[Kee](#)-[Dirac](#)-[Narayanan](#). Ground 2 adds [Kamruzzaman](#) for Claim 4; Ground 3 adds [Wang](#) for Claim 5; Ground 4 adds [Kamruzzaman](#) and

removes **Kee**⁵ for Claims 7-11 and 13; Ground 5 adds **Wang** and removes **Kee** for Claims 14-17 and 19-20; and Ground 6 adds **Wang** and **Kamruzzaman** while removing **Kee** for Claims 12 and 18.

Because Petitioner has not shown that Grounds 1, 4, and 5 satisfy independent Claims 1, 7, and 14, Petitioner cannot show that the dependent claims are unpatentable in Grounds 2, 3, and 6. The additional references are used only for additional dependent limitations. They do **not** cure Ground 1's failure to disclose a coherent combined system **that combines first-user accent features with second-user natural-voice characteristics to generate output audio that preserves aspects of the second user's voice while mimicking the first user's accent**. The Board should deny Grounds 2, 3, and 6 because they depend on Grounds 1, 4, and 5 respectively.

⁵ **Kee** is removed for all claims not dependent on Claim 1 because Claim 1 is the only claim that requires an audio interface.

VI. CONCLUSION

Ground 1 is the foundation of the Petition, and Petitioner does not show a supported motivation to combine Nguyen-**Kee-Dirac-Narayanan** or a reasonable expectation of success. Because Grounds 2-6 depend on Nguyen-**Dirac-Narayanan**'s defective base combination, those grounds fail as well. Furthermore, Grounds 2-6 each fail because of their own defective combinations.

For the foregoing reasons, Patent Owner respectfully requests that the Board deny institution of post-grant review of Claims 1-20 of U.S. Patent No. 12,417,756.

Respectfully submitted,

AMSTER, ROTHSTEIN & EBENSTEIN LLP
Attorneys for Patent Owner
405 Lexington Avenue, 48th Floor
New York, NY 10174
(212)336-8000

Dated: June 30, 2026
New York, New York

By: /Charles R. Macedo/
Charles R. Macedo (Reg. No. 32,781)

Of counsel:
Gary J. Gershik (Reg. No. 39,992)
Lewis Derenzis III (Reg. No. 84,132)

CERTIFICATE OF WORD COUNT

The undersigned certifies that the foregoing PATENT OWNER'S PRELIMINARY RESPONSE complies with the type-volume limitation in 37 C.F.R. § 42.24(b)(1). According to the word-processing system's word count, the brief contains approximately 11,601 words, excluding the parts exempted by 37 C.F.R. § 42.24(a).

Respectfully submitted,

AMSTER, ROTHSTEIN & EBENSTEIN LLP
Attorneys for Patent Owner
405 Lexington Avenue, 48th Floor
New York, NY 10174
(212)336-8000

Dated: June 30, 2026
New York, New York

By: /Charles R. Macedo/
Charles R. Macedo (Reg. No. 32,781)

CERTIFICATE OF SERVICE

I hereby certify that on June 30, 2026, a true and correct copy of the foregoing PATENT OWNER'S PRELIMINARY RESPONSE was served via email, by consent, to Petitioner's counsel of record at the correspondence addresses of record in this proceeding.

Eric A. Buresh (Reg. No. 50,394)
Chris R. Schmidt (Reg. No. 63,982)
Nick R. Apel (Reg. No. 84,017)
Adam M. Sandwell (Reg. No. 72,484)
Erise IP, P.A.
7015 College Blvd., Suite 700
Overland Park, Kansas 66211
Telephone: (913) 777-5600
Fax: (913) 777-5601
chris.schmidt@eriseip.com
nick.apel@eriseip.com
adam.sandwell@eriseip.com
PTAB@eriseip.com

Respectfully submitted,

AMSTER, ROTHSTEIN & EBENSTEIN LLP
Attorneys for Patent Owner
405 Lexington Avenue, 48th Floor
New York, NY 10174
(212)336-8000

Dated: June 30, 2026
New York, New York

By: /Charles R. Macedo/
Charles R. Macedo (Reg. No. 32,781)