

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

KRISP TECHNOLOGIES, INC.,
Petitioner,

v.

SANAS.AI INC.,
Patent Owner.

Case No. PGR2026-00032
Patent No. 12,412,561

PATENT OWNER'S PRELIMINARY RESPONSE

Table of Contents

I.	INTRODUCTION	1
I.	THE '561 PATENT	7
II.	CLAIM CONSTRUCTION.....	10
III.	THE PETITION FAILS TO DEMONSTRATE MOTIVATION, REASONABLE EXPECTATION OF SUCCESS, AND PARTICULARIZED MAPPING OF CLAIM ELEMENTS.....	12
	A. The Petition Does Not Demonstrate a Reasoned Motivation to Combine Prior Art References	21
	B. The Petition Does Not Demonstrate a Reasonable Expectation of Success	25
	C. The Petition Does Not Make a Particularized Showing of Where Each Claim Element is Disclosed in the Asserted Prior Art.....	28
	D. Institution Should Be Denied as an Inefficient Use of Office Resources	31
IV.	PETITIONER'S CHALLENGES FAIL ON THE MERITS.....	32
	A. Ground 1 Fails on the Merits	33
	1. Ground 1 Independently Fails Because Prabhavalkar is Not Analogous Art.....	34
	2. Petitioner Fails to Show the Examiner-Identified Inventive Concept of the Spectrum Representation and the Spectrum Representation's use in the Vocoder	36
	3. Petitioner Fails to Show Extracting Linguistic Features with a Reduced Accent	41
	4. Petitioner Fails to Show Speaker-Embedding Integration ..	43
	5. Petitioner Improperly Treats Fan's Distinct Embodiments as Interchangeable.....	46
	6. Petitioner Fails to Show Context-Based Feature Extraction	49
	7. Dependent Claim Challenges of Ground 1 Do Not Cure the Independent-Claim Defects	52
	B. Grounds 2-4 Fail on the Merits	53
	1. Grounds 2-4 Fail Because They Rely on Ground 1	53
	2. Ground 4 Also Fails Because Watts Does Not Teach the Claimed Time-Shift Parameter	53

3.	Ground 3 Also Fails Because Kalinli’s Parallel Acoustic Filters Do Not Teach the Claimed Parallel Processing	55
4.	Ground 2 Also Fails Because Sargsyan Does Not Cure the Missing Claimed Pipeline.....	57
V.	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)	21
<i>Apple Inc. v. Ginko LLC</i> IPR2025-01388, Paper 14 (March 2, 2026)	30
<i>In re Bigio</i> , 381 F.3d 1320 (Fed. Cir. 2004)	34
<i>Chevron Oronite Co. v. Infineum USA</i> , IPR2018-00923, Paper 9 (November 7, 2018)	31
<i>CyberSecure IPS, LLC v. Network Integrity Systems, Inc.</i> , IPR2025-01441, Paper 18 (May 26, 2026)	31
<i>Honeywell Int'l Inc. v. 3G Licensing, S.A.</i> , 124 F.4th 1345 (Fed. Cir. 2025)	26
<i>Intelligent Bio-Sys., Inc. v. Illumina Cambridge Ltd.</i> , 821 F.3d 1359 (Fed. Cir. 2016)	25
<i>In re Kahn</i> , 441 F.3d 977 (Fed. Cir. 2006)	21
<i>KSR Int'l Co. v. Teleflex Inc.</i> , 550 U.S. 398 (2007)	12, 21
<i>PAR Pharm., Inc. v. TWI Pharms., Inc.</i> , 773 F.3d 1186 (Fed. Cir. 2014)	25
<i>Virtek Vision Int'l ULC v. Assembly Guidance Sys., Inc.</i> , 97 F.4th 882 (Fed. Cir. 2024)	45
<i>Xerox Corp. v. Bytemark, Inc.</i> , IPR2022-00624, Paper 9 (PTAB Aug. 24, 2022)	21

STATUTES

35 U.S.C. § 312(a)(3).....	<i>passim</i>
35 U.S.C. § 314(a)	12
37 C.F.R. § 42.104(b)(4).....	28
37 C.F.R. § 42.207	1

PATENT OWNER’S EXHIBIT LIST

<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.) (“ <i>Sanas.AI Inc. v. Krisp Technologies, Inc.</i> 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, Sanas</i> , 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)

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 1–18 of U.S. Patent No. 12,412,561 (the “’561 Patent”).

Institution should be denied on the merits.

The ’561 Patent claims an **end-to-end accent correction pipeline – dual context-based extraction, spectrum representation synthesis, and output generation**. The Examiner understood exactly that and allowed the claims because no prior art taught or suggested “synthesizing a spectrum representation based on the acoustic features, the linguistic features, and a speaker embedding for a human speaker”. Ex1002, at 176-177.

Petitioner never identifies any reference—or combination—that discloses or renders obvious this allowance-dependent limitation in the context of the claimed pipeline. Its underlying theory of prior art combinations is illogical, cobbling together discrete fragments from three disconnected references directed to different problems and architectures, as illustrated by lynchpin Claim 1:

[1(Pre)] A method for real-time accent correction, the method implemented by a computing system and comprising:

[1(a)] extracting, from a chunk of a stream of chunks of an input audio signal, acoustic features and linguistic features, wherein the linguistic features have a reduced accent;

[1(b)] synthesizing a spectrum representation based on the acoustic features, the linguistic features, and a speaker embedding for a human speaker; and

[1(c)] generating, by a vocoder and based on the spectrum representation, an output chunk of an output audio signal.

Ex1001, Cl. 1 (color coding supplied). The Petition fails to address—let alone map with particularity—the very limitation the Examiner identified as warranting allowance: synthesizing a spectrum representation based on the acoustic features, the linguistic features, and a speaker embedding for a human speaker. Ex1002, at 176-177.¹

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

¹ This is not the only time Petitioner has advanced this same defective theory. The Petition repeats the same failed Feinauer–**Prabhavalkar**–Fan combination that Petitioner presented against the ’561 Patent’s parent patent in IPR2026-00275. It merely repackaged the combination for the claims here. That repetition underscores the hindsight-driven nature of the challenge: Petitioner again attempts to reconstruct the claimed real-time accent-correction architecture from isolated fragments of unrelated systems, without showing that the asserted references disclose or would have motivated the claimed integrated pipeline.

- (1) Feinauer (shown in black text) discloses *a context-driven accent and dialect modification framework* that operates using translation models, dictionaries, pairing databases, machine-learning models, and context selection to modify speech characteristics. Ex1004 at ¶¶[0047]–[0057], [0061]–[0078], [0104]–[0117]; the petition *does not assert that it discloses the claimed pipeline or, critically, any extraction of linguistic features with reduced accent from chunk-plus-context as required by the claims, let alone output any signal using these extracted features and speaker embedding*;
- (2) **Prabhavalkar (shown in blue bold text)**, by contrast, **is not directed to accent modification at all—it is an automatic speech recognition (“ASR”) system that processes speech frames through a sequence-to-sequence encoder-attender-decoder architecture to generate transcriptions**, and **does not generate accent-modified audio or any linguistic feature with a reduced accent**. Ex1005, Abstract, 1:15–25, 2:41–3:56, 6:56–7:12, 11:19–57. While **Prabhavalkar is being offered to show the use of chunking**, it is a totally different and irrelevant context—**Prabhavalkar is non-analogous art**;
- (3) Fan (shown in green underscore text), in turn, discloses voice-conversion architectures with selectable modes—such as style conversion, dialect conversion, and voice enhancement—**each implemented through distinct networks, feature representations, and model pathways for synthesizing output speech**. Ex1006 at ¶¶[0005]–[0007], [0045], [0055]–[0060], [0084]–[0095], [0016]–[0019]. But Fan does not disclose extracting linguistic features with a reduced accent from a chunk and its context in the manner required by the claims.

That color-coded patchwork of Feinauer, **Prabhavalkar**, and Fan 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 limitation, which is nowhere disclosed 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 to generate an accent-corrected output audio signal based on acoustic features, linguistic features, and a speaker embedding. Instead, it purportedly reconstructs the claims from independent references without showing how the combination would yield a system that actually performs the claimed method.

Petitioner's reliance on this pieced-together combination is a de facto admission that **no single reference—or any coherent combination—teaches the claimed operation. Nor does any reference—or combination—teach the Examiner-identified requirement of synthesizing a spectrum representation based on the acoustic features, the linguistic features, and a speaker embedding for a human speaker, which remains the critical missing piece in**

the Petition’s theory. Each reference addresses different data, objectives, and architectures, and the Petition does not explain how those fundamentally different systems could be combined to perform the claimed real-time accent correction pipeline, which utilizes chunking, acoustic and linguistic feature extraction, and speaker embedding integration to generate a spectrum representation to generate an accent corrected audio signal.

Petitioner’s theory throws together an accent and dialect modification framework/system (Feinauer), a **chunking format (Prabhavalkar)**, and a separate voice-conversion module (Fan), and then declares the result a real-time accent-correction engine. Those components may all fall under the broad umbrella of speech processing, but they operate on different data, pursue different objectives, and produce different outputs. Their superficial similarity does not amount to a coherent disclosure of the claimed method.

As the Petitioner does not identify any prior-art teaching of the claimed arrangement and operation of a real-time accent-correction 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 precise feature that supported allowance: synthesizing a spectrum representation based on

the acoustic features, the linguistic features, and a speaker embedding for a human speaker. See Ex1002, at 176-77.

The remaining independent claim challenges similarly fail at least on these same defects. Claims 1, 6, 11, and 16-18 each require the same core integrated pipeline. See Ex1001 at Cl. 1, 6, 11, and 16-18. The Examiner allowed the claims because the prior art failed to teach that integrated synthesis sequence, and the defective Claim 1 mapping equally applies to the remaining independent claims. Ex1002 at 660. Because the same deficiencies infect the Petition's inadequate reading on all six independent claims, institution must be denied.

Grounds 2-4 offer no cure for the flaws of Ground 1. They depend on the same deficient Feinauer-[Prabhavalkar-Fan](#) base combination and merely pile on even more disparate and irrelevant references that have no teachings to cure this defect, see Sargsyan (Ex1007), Kalinli (Ex1009), and Watts (Ex1008) purportedly for dependent-claim details. Putting aside whether they separately state a claim for the other grounds, for present purposes what matters is that none are offered to supply the missing Examiner-identified limitations or convert Petitioner's disconnected fragments into a coherent real-time accent-conversion pipeline.

This failure is dispositive: the Petition **does not** satisfy its burden under 35 U.S.C. § 312(a)(3) to show **with particularity where each claim element is**

found in the prior art, instead reconstructing the claims from disconnected fragments.

II. THE '561 PATENT

The '561 Patent is directed to “real-time correction of an accent in an input audio signal.” Ex1001 at Abstract. Claim 1 recites a specific pipeline for real-time accent correction which extracts acoustic features and linguistic features from chunks to generate an output based on these features and speaker embeddings. *Id.* at Cl. 1.

Independent claims 1, 6 and 11 operate by continuously segmenting an incoming speech audio signal into chunks and processing each chunk in real time. Two different types of features are extracted from the input chunks: acoustic features of the speaker’s voice, such as pitch, energy, and whether speech is present; and linguistic features having a reduced accent, which capture what the speaker actually said. These features and a speaker embedding are used to synthesize a spectrum representation of the speaker’s voice, which is then used generate an output audio signal.

Independent Claims 6 and 11 recite corresponding system and computer-readable-medium implementations of claim 1. Independent Claims 16–18 recite related independent method, system, and computer-readable-medium implementations that insert claim limitations, with the primary variation being the

requirements of (1) dividing a speech audio signal into a stream of input chunks, an input chunk from the stream of input chunks including a pre-defined number of frames of the speech audio signal; and (2) obtaining a speaker embedding for a human speaker.

Claim 1's method requires:

[1(Pre)] A method for real-time accent correction, the method implemented by a computing system and comprising:

[1(a)] extracting, from a chunk of a stream of chunks of an input audio signal, acoustic features and linguistic features, wherein the linguistic features have a reduced accent;

[1(b)] synthesizing a spectrum representation based on the acoustic features, the linguistic features, and a speaker embedding for a human speaker; and

[1(c)] generating, by a vocoder and based on the spectrum representation, an output chunk of an output audio signal.

Ex1001, Cl. 1 (color coding supplied).

As labeled in the Introduction, these elements define **a single, integrated real-time pipeline that must operate cohesively** by performing dual context-based extraction, integrating a speaker embedding, and generating outputs.

Critically, element **[1(b)]**—“synthesizing a spectrum representation based on the acoustic features, the linguistic features, and a speaker embedding for a human speaker”—was recognized by the Examiner as not taught or suggested by the prior

art and as a basis for allowance. Ex1002, at 176-177.² This was not a peripheral observation; it reflects the Office's determination that the prior art failed to disclose this key feature that differentiates the claimed real-time pipeline.

Instead, the Petition simply ignores the context of the allowance-dependent limitation while 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. § 312(a)(3)** to show where this critical element is disclosed.

The Petition:

- **does not show** any teaching or suggestion in the asserted art of the integrated sequence of chunk-based extraction and generation;
- **does not sufficiently explain how** Prabhavalkar's chunk-based format or Fan's style embedding vector could be incorporated into Feinauer's accent and dialect modification framework/system; and

² The Office recognized that this novel and inventive pipeline applied to all the pending claims, including the other independent claims 6, 11, and 16-18. Ex1002, at 176-177. In a similar manner, the Petition applies the same defective analysis to all three independent claims. Pet. at 33-35, and 37-43.

- ***improperly treats*** Fan's style separation and dialect conversion embodiments as interchangeable.

Significantly, the Petition fails to demonstrate that even if this illogical combination were made that it would result in the claimed invention. By failing to identify where the allowance-dependent limitations are disclosed in the prior art, and by failing to explain **how** the proposed combination would perform those limitations as an integrated whole, the Petition lacks the particularized showing required by 35 U.S.C. § 312(a)(3) and should be rejected.

III. CLAIM CONSTRUCTION

Petitioner asserts that the claims should be interpreted under the *Phillips* standard and applies the plain and ordinary meaning of the claim terms as understood by a PHOSITA. Pet. at 3. For purposes of institution, Patent Owner agrees that no express claim construction is necessary because Petitioner's grounds fail under the plain and ordinary meaning of the claims. The institution determination on the merits requires the Petition to demonstrate that the asserted prior art teaches "extracting ... linguistic features, wherein the linguistic features have a reduced accent", and the Petition does not come close to demonstrating that this element is present in the art asserted here. Accordingly, there is no claim

construction dispute material to the institution decision. Patent Owner reserves the right to propose and address claim constructions should trial be instituted.

IV. THE PETITION FAILS TO DEMONSTRATE MOTIVATION, REASONABLE EXPECTATION OF SUCCESS, AND PARTICULARIZED MAPPING OF CLAIM ELEMENTS

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, 312(a)(3), 314(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. Each independently show that there is not a reasonable likelihood of success that the petitioner would prevail. 35 U.S.C. § 314(a).

Here, the Petition fails because it **never shows the claimed integrated real-time accent-correction pipeline**. The reason the Petition cannot make this showing is that the alleged Feinauer–**Prabhavalkar**–**Fan** combination is inoperable at each critical step. Petitioner does not identify a coherent prior-art architecture that performs the claimed method. Instead, it attempts to cobble together Feinauer’s high-level accent/dialect modification framework, **Prabhavalkar**’s ASR-specific chunking and temporal context, and **Fan**’s neural

voice-conversion components. But no amount of additional explanation can reconcile the architectural defects in that proposed combination.

For the chunk-based portion of Claim [1(a)], the proposed combination of **Prabhavalkar**'s ASR chunk format with Feinauer's accent/dialect modification framework is not functional:

- Feinauer is directed to modifying accents and dialects in voice communications using an accent and dialect modifier, context selector, translation models, dictionaries, and machine-learning models; and
- **Prabhavalkar**, by contrast, is an ASR reference that discloses “a method for performing speech recognition” that “generat[es] a transcription... provided as an output of the ASR system.” Ex1005, Abstract. Prabhavalkar's chunking is tied to an ASR architecture, **not a speech-modification pipeline**. It explains that because full-sequence LAS requires “the entire utterance” before decoding, a “‘chunk’-based approach” may process attention over “small segments of the utterance, such as a certain number of frames,” shifting attention between chunks so the system can output a transcript

with lower latency—not generate accent-modified audio. Ex1005, 17:33-51.

Thus, **Prabhavalkar**'s chunks are not a deployable processing component for Feinauer's accent-modification architecture.

The Petition **never explains** how Feinauer's word-level accent/dialect modifier would be restructured to receive and process **Prabhavalkar**-style ASR chunks, or why a PHOSITA would expect ASR chunks to operate within Feinauer's non-ASR modification framework. The proposed Feinauer–**Prabhavalkar** combination therefore cannot coherently perform the claimed extraction “from a chunk of a stream of chunks of an input audio signal.”

For the acoustic-feature extraction portion of Claim [1(a)], the proposed combination also fails. Claim 1 requires extracting acoustic features from the claimed chunk of the input audio signal. Petitioner relies on:

- Feinauer for general prosodic characteristics;
- Fan for feature-extraction architecture; and
- **Prabhavalkar** for chunking.

The operation of the proposed prior art references is critical in their failure to render the claimed pipeline:

- **Prabhavalkar**'s chunking and attention mechanisms are part of an ASR encoder-attender-decoder architecture used to support speech recognition, not accent-corrected speech generation; and
- **Fan**'s **feature-extraction models operate within voice-conversion embodiments for selected conversion modes**: Fan discloses that the terminal selects among "a style conversion mode, a dialect conversion mode, and a voice enhancement mode," then extracts content using a "voice feature extraction model" that "extracts a phoneme posteriorgram PPG feature," while the style-conversion path uses a "style separation model" and sub-feature models to extract style-related features including "timbre," "rhythm," "accent," and "speaking speed." Ex1006 ¶¶[0005], [0007]–[0009].
- Petitioner never explains **how** **Prabhavalkar**'s ASR chunking would be incorporated into **Fan**'s feature-extraction architecture, **or how** that modified **Fan** component would then be grafted into Feinauer's accent/dialect modifier. Without that explanation, the Petition does not show a working acoustic-feature extraction module that extracts acoustic features from the claimed input chunk.

For the “linguistic features ... with a reduced accent” portion of Claim [1(a)], the proposed combination is likewise deficient. Claim 1 requires extracting linguistic features from the chunk, where those linguistic features have a reduced accent. Petitioner points to [Fan](#)’s PPG-related disclosure for linguistic features, while relying on [Prabhavalkar](#) for chunking. But [Fan](#)’s PPG extraction is not shown to operate on [Prabhavalkar](#)-style chunks, and [Prabhavalkar](#)’s chunking is designed for ASR decoding rather than reduced-accent linguistic feature extraction.

Claims 16–18 expressly require extraction “from the input chunk and a context associated with the input chunk.” For this limitation, Petitioner relies on [Prabhavalkar](#)’s prior-chunk ASR context and [Fan](#)’s feature-extraction disclosures. But [Prabhavalkar](#)’s context is used by an ASR model to compute attention and generate recognition outputs, while [Fan](#)’s PPG and style-feature extraction pathways are not shown to operate on an input chunk and associated context. The Petition therefore does not identify a feature-extraction architecture that extracts either acoustic features or reduced-accent linguistic features from both the input chunk and its associated context, as Claims 16–18 require.

For the spectrum-representation synthesis step of Claim [1(b)], the proposed combination again fails. Claim 1 requires “synthesizing a spectrum representation based on the acoustic features, the linguistic features, and a speaker embedding for a human speaker.” Petitioner relies principally on [Fan](#)’s neural voice-conversion architecture for this limitation. But [Fan](#)’s melspectrogram or mel feature is generated within [Fan](#)’s own conversion network, using [Fan](#)’s own feature representations and network pathways. **Feinauer does not disclose a PPG-to-spectrum-to-vocoder pipeline like [Fan](#)** (and thus does not fit); it discloses an accent/dialect modification framework based on accent/dialect detection, translation models, dictionaries, and contextual selection.

Petitioner never explains **how** [Fan](#)’s synthesis architecture would be inserted into Feinauer’s framework, **what** components of Feinauer would be retained, **what** components of [Fan](#) would replace them, or **how** the resulting system would synthesize a spectrum representation based on all three claimed inputs.

For the speaker-embedding aspect of Claim [1(b)], Petitioner’s combination is also inoperable. Petitioner relies on [Fan](#)’s style embedding vector as the claimed speaker embedding, while invoking Feinauer only as a high-level accent-modification backdrop. **But [Fan](#)’s style embedding vector is not a free-floating value that can simply be inserted into Feinauer.** It is generated through [Fan](#)’s

own neural voice-conversion architecture, including style separation, sub-feature extraction, user-selected weights, and multihead attention. Ex1006, ¶¶ [0087]-[0095].

Feinauer has no corresponding neural synthesis architecture for receiving that embedding, combining it with acoustic and linguistic features, and generating output speech. Petitioner never explains how [Fan](#)'s embedding-generating neural network would interface with Feinauer's accent/dialect modifier, which components would be retained or replaced, or what data representation would be passed between them. The alleged combination therefore does not supply an operable speaker-embedding architecture.

The proposed combination also fails at the vocoder/output-generation step of Claim **[1(c)]**. Claim 1 requires "generating, by a vocoder and based on the spectrum representation, an output chunk of an output audio signal." Petitioner attempts to satisfy this limitation by combining [Prabhavalkar](#)'s chunk-based processing, [Fan](#)'s neural voice-conversion/vocoder architecture, and Feinauer's general accent/dialect modification framework. But those components do not fit together:

- [Prabhavalkar](#)'s chunk outputs are ASR recognition outputs, not output audio chunks;

- [Fan](#) can generate speech through a neural voice-conversion pathway and vocoder, but that pathway depends on [Fan](#)'s own feature representations and network structure; and
- Feinauer does not disclose a PPG-to-spectrum-to-vocoder pipeline like Fan (and thus does not fit); it modifies voice communications through its accent/dialect modification framework.

The Petition never provides an operable way to integrate [Prabhavalkar](#)'s chunks, [Fan](#)'s neural synthesis components, and Feinauer's accent/dialect modifier into a system that generates output audio chunks based on the claimed acoustic features, reduced-accent linguistic features, and speaker embedding.

The further cobbling together of [Fan](#)'s distinct embodiments **only compounds the problem**. [Fan](#) discloses separate selectable modes, including style conversion, dialect conversion, and voice enhancement, each with corresponding networks and pathways. Petitioner **treats those embodiments as interchangeable**, relying on [Fan](#)'s style-conversion components for embeddings and acoustic/style features while also invoking [Fan](#)'s dialect-conversion pathway for accent modification. But the Petition does not explain **why** a PHOSITA would combine those distinct [Fan](#) embodiments with each other, much less **how** that internally reconstructed [Fan](#) system would then be integrated into Feinauer and

Prabhavalkar. This is not routine reprogramming; it is an **unsupported reconstruction** of the '561 Patent's claimed architecture using isolated fragments from unrelated systems, resulting in an **inoperable combination**.

Accordingly, the alleged Feinauer–**Prabhavalkar**–Fan combination does not teach or render obvious the claimed integrated pipeline. Feinauer supplies only a high-level accent/dialect modification framework, **Prabhavalkar** supplies ASR-specific chunking and context, and Fan supplies neural voice-conversion components from distinct embodiments. The Petition does not provide an operable path for those components to work together to perform the claimed sequence of steps. 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 separately renders the proposed combination non-operable as the claimed pipeline. The result is not an integrated system that extracts, synthesizes, and generates as the claims require, but a set of disconnected components that cannot be assembled into the claimed invention without rebuilding the system. These failures are carried through all the grounds which rely upon this defective analysis.

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

First, 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”, “reprogramming”, or “predictable results”—are insufficient where, as here (*see e.g.* Pet. at 17, 20, 23, 25, 31, 37, 39, 48, 51, 55, 60, 63, 67-68, 71, 74), 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 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”).

Here, as discussed in Section V(A), Petitioner **does not provide a reasoned explanation** for how Feinauer, [Prabhavalkar](#), and [Fan](#) would be combined into a coherent system that performs the claimed real-time accent-correction pipeline. While the Petition identifies purported reasons why a PHOSITA would have combined those references, those assertions are conclusory and fail to address the technical incompatibilities noted above that such a combination entails.

With respect to [Prabhavalkar](#), the Petition argues that Feinauer’s high-level disclosure of processing speech “portions” could be implemented using [Prabhavalkar](#)’s ASR chunking, supposedly through routine reprogramming. Pet. at 13–15. But the Petition never explains how [Prabhavalkar](#)’s chunking and temporal-context framework could function within Feinauer’s accent/dialect modification architecture. **That incompatibility is fatal:**

- Feinauer modifies voice communications using an accent and dialect modifier, translation models, dictionaries, pairing databases, machine-learning models, and context-based selection; and
- [Prabhavalkar](#), by contrast, uses chunks in an ASR system that processes speech frames through a sequence-to-sequence encoder-attender-decoder

architecture to generate recognition outputs, such as word elements or transcriptions—not modified speech audio.

The Petition **offers only the unsupported assertion** that [Prabhavalkar](#)'s ASR chunking can be incorporated into Feinauer's non-ASR accent-modification system, without sufficient detail, logic, or evidentiary support to justify institution.

The same defect appears in the Petition's reliance on [Fan](#). [Fan](#) supplies the specific neural voice-conversion components that Petitioner needs for feature extraction, embeddings, spectrum generation, and vocoder-based output. **But [Fan](#)'s architecture is not a plug-in component for Feinauer.**

[Fan](#) extracts PPGs and style features within distinct voice-conversion embodiments, including style conversion, dialect conversion, and voice enhancement, each with its own corresponding network and model pathway. The Petition treats those distinct [Fan](#) embodiments as interchangeable, asserting without meaningful support or analysis that [Fan](#)'s style-separation and embedding components would be equally applicable to [Fan](#)'s dialect-conversion pathway and then further insertable into Feinauer. Pet. at 16–31. But the Petition never explains how [Fan](#)'s neural feature representations, style embedding vector, synthesis pathway, or vocoder would be inserted into Feinauer's accent/dialect modification framework. Thus, the **Petition fails to explain not only how the proposed**

combination would produce the claimed result, but also why a PHOSITA would have been motivated to make that combination in the first instance.

The Petition's treatment of [Fan](#)'s speaker embedding highlights the problem. Feinauer provides, at most, a high-level backdrop for modifying speech characteristics, while Petitioner's actual speaker-embedding architecture is imported from [Fan](#). [Fan](#)'s style embedding vector is generated through [Fan](#)'s own neural architecture, including style separation, sub-feature extraction, user-selected weights, and multihead attention. Feinauer, however, **does not disclose a corresponding neural synthesis pipeline** capable of receiving that embedding, combining it with acoustic and linguistic features to synthesize a spectrum representation, and producing output audio through a vocoder. The Petition never explains what component in Feinauer would receive [Fan](#)'s embedding, what data representation would be passed between the systems, or how [Fan](#)'s embedding-generating model would interface with Feinauer's accent/dialect modifier.

The Petition's reliance on [Fan](#)'s spectrum-representation/vocoder architecture is similarly unsupported. Claim 1 requires synthesizing a spectrum representation based on the acoustic features, the linguistic features, and a speaker embedding, and then generating an output chunk using a vocoder based on that spectrum representation. [Fan](#)'s vocoder operates within [Fan](#)'s own voice-

conversion pathway, where a mel feature or mel-spectrum representation is produced by Fan's conversion model and then converted into playable voice by the neural network vocoder. Feinauer does not disclose such a PPG-to-spectrum-to-vocoder architecture like Fan (and thus does not fit). Instead, Feinauer discloses real-time accent/dialect modification using accent/dialect detection and modification models. The Petition never explains **why a PHOSITA would redesign** Feinauer to generate Fan-style spectrum representations merely so that Fan's vocoder could be used, or **how such a redesigned system would preserve the claimed relationship** among acoustic features, reduced-accent linguistic features, speaker embedding, spectrum representation, and output audio chunk.

Because the Petition relies on conclusory assertions of convenience and ease of modification rather than reasoned analysis addressing the incompatibility, it fails to establish the required motivation to combine.

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

Second, 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, 1355 (Fed. Cir. 2025) (quotation omitted).

Here, as discussed in Section V.A, a PHOSITA would have no adequate basis to expect that the proposed Feinauer–[Prabhavalkar](#)–[Fan](#) combination would succeed in producing the integrated real-time accent-correction pipeline that the challenged independent claims require.

Nor does the Petition explain how these fundamentally different architectures could be integrated to perform the claimed operations, particularly the allowance-dependent limitation that was found missing from the prior art:

- Feinauer is a context-driven accent and dialect modification system that uses an accent and dialect modifier, translation models, dictionaries, pairing databases, machine-learning models, and context selection to modify speech characteristics;
- [Prabhavalkar](#) is an ASR system that processes speech frames through a sequence-to-sequence encoder-attender-decoder architecture

to generate recognition outputs, such as word elements or transcriptions—not modified speech audio; and

- [Fan](#) is a neural voice-conversion system with distinct selectable modes—such as style conversion, dialect conversion, and voice enhancement—implemented through separate model pathways, including PPG extraction, style separation, embedding generation, mel-feature generation, and vocoder-based output.

These are not merely different references but **materially different architectures**, and the Petition **never explains why a skilled artisan would have expected** Feinauer’s word-level accent/dialect modification framework, [Prabhavalkar](#)’s ASR-specific chunk-based format and attention-based context, and [Fan](#)’s neural PPG/style-embedding/synthesis/vocoder pipeline to interoperate within a single end-to-end real-time accent-correction system. A combination asserted at the broad level of generality of the Petition, untethered from any substantive analysis bridging these architectural differences, demonstrates no expectation of success at all, and cannot sustain the obviousness ground.

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

Third, 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. § 312(a)(3); 37 C.F.R. § 42.104(b)(4).

Here, the Petition maps individual elements of independent Claims 1, 6, 11, and 16–18 to scattered disclosures across three references—Feinauer for the general accent/dialect modification framework, [Prabhavalkar](#) for chunk-based format and temporal context, and [Fan](#) for feature extraction, speaker embeddings, spectrum representation, and vocoder-based output—but never makes a particularized showing that those pieces combine into the claimed pipeline as an operable whole. Pet. at 11–43. That failure is especially clear for Claim **[1(a)]–[1(c)]**. Those limitations do not merely require locating some speech-processing concepts in the prior art; they require an **integrated real-time accent-correction pipeline** that extracts acoustic features and reduced-accent linguistic features from chunks, synthesizes a spectrum representation based on those features and a speaker embedding, and generates an output audio chunk by a vocoder.

Neither [Prabhavalkar](#) nor [Fan](#) teaches that integrated limitation:

- **Prabhavalkar** provides ASR-specific chunking and context for a recognition system that generates word elements or transcriptions, not modified speech audio; and
- **Fan**, by contrast, extracts PPGs and style/acoustic features within neural voice-conversion embodiments.

The Petition does not show **how** **Fan** could extract those features from **Prabhavalkar**-style input chunks, or—where Claims 16–18 are concerned—from input chunks together with prior-chunk context. Thus, pointing to **Prabhavalkar** for chunking/context and **Fan** for PPG or style extraction is not sufficient. The Petition needed to explain **how** **Prabhavalkar**'s ASR temporal-context mechanism could be integrated into **Fan**'s feature-extraction architecture and then incorporated into Feinauer's accent/dialect modification framework. **The Petition provides no such explanation.**

That omission is fatal under § 312(a)(3). The claims require the chunking, acoustic-feature extraction, reduced-accent linguistic-feature extraction, speaker embedding, spectrum-representation synthesis, and output-audio-chunk generation steps to operate **within a single integrated system**. The Petition **never supplies that missing integration**. It instead identifies isolated parts in separate references and assumes the claimed system follows.

Identifying where each limitation appears in some reference, standing alone, does not discharge Petitioner's burden. Claim 1 requires that the extraction, speaker-embedding, synthesis, and output-generation steps operate together in a real-time accent-correction pipeline. Claims 16–18 further require dividing the speech audio signal into input chunks and extracting acoustic and linguistic features from an input chunk and associated context. **The Petition offers no particularized analysis for that integrated operation.**

Instead, Petitioner **cobbles the claimed limitations** together from separate and disparate sources **without showing**, element by element, *how* Feinauer's accent/dialect modification framework, [Prabhavalkar's](#) ASR chunk-based format and temporal context, and [Fan's](#) neural feature-extraction and voice-conversion components combine to form a coherent pipeline, **let alone that such a combination would result in the claimed accent-correction pipeline.**

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 (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. Petitioner's showing here falls short for the same reasons: it neither identifies a prior-art disclosure of the **claimed**

chunk-based extraction of reduced-accent linguistic features, spectrum-representation synthesis, and output-audio-chunk generation within an integrated system, nor explains how the proposed Feinauer–[Prabhavalkar](#)–[Fan](#) combination performs **those requirements within the claimed pipeline**.

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

Finally, denial is warranted where 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 (May 26, 2026) (citing *Chevron Oronite Co. v. Infineum USA*, IPR2018-00923, Paper 9 at 10–11 (November 7, 2018) (informative)).

Here, all asserted grounds depend on the same defective Feinauer, [Prabhavalkar](#), and [Fan](#) combination, and the additional references in Grounds 2-4 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. PETITIONER’S CHALLENGES FAIL ON THE MERITS

Petitioner asserts four obviousness grounds:

Ground	Claims Challenged	References Relied Upon	Petition Citation
Ground 1	1–2, 6–7, 11–12, 16–18	Feinauer, Prabhavalkar , Fan	Pet. 11–43
Ground 2	3, 8, 13	Feinauer, Prabhavalkar , Fan , Sargsyan	Pet. 43–47
Ground 3	4, 9, 14	Feinauer, Prabhavalkar , Fan , Kalinli	Pet. 47–50
Ground 4	5, 10, 15	Feinauer, Prabhavalkar , Fan , Watts	Pet. 50–53

This Preliminary Response focuses on Ground 1 because Ground 1 supplies the base combination that every other asserted ground, and every claim, requires. Ground 1 asserts that claims 1–2, 6–7, 11–12, and 16–18 are obvious over Feinauer, [Prabhavalkar](#), and [Fan](#). Pet. at 11. Grounds 2–4 merely add Sargsyan, Kalinli, or Watts for dependent-claim limitations, while retaining the same

Feinauer-[Prabhavalkar](#)-[Fan](#) combination for the independent-claim limitations.

Pet. at 43, 47, 50. Thus, if Ground 1 fails to show the claimed alignment in independent claims 1, 6, 11, and 16–18, Grounds 2–4 fail as well.

A. Ground 1 Fails on the Merits

Ground 1 fails because Petitioner attempts to reconstruct the '561 Patent through hindsight rather than identify a teaching or reasoned combination that a PHOSITA at the time of the invention would have recognized in the prior art. The Petition relies on three references that perform meaningfully different functions using materially distinct architectures. Feinauer modifies speech using an accent and dialect modifier system. Ex1004 at ¶[0047]. [Prabhavalkar](#) recognizes speech using ASR architecture that outputs transcriptions. Ex1005 at Abstract, 1:15–25, 2:41–3:56. [Fan](#) converts speech based on the selection of a particular voice-conversion mode. Ex1006 at ¶[0005].

The Petition never bridges those differences. It does **not**:

- identify the combined architecture,
- explain which components are retained or discarded, or
- show how outputs from one system become usable inputs to another.

In particular, the Petition does **not** explain:

- **how** Feinauer’s accent modifier model would process **Prabhavalkar**’s ASR chunks,
- **how** **Fan**’s PPG extractor would receive **Prabhavalkar**’s previous-chunk context, or
- **how** **Fan**’s embedding-based neural conversion pathway would be integrated into Feinauer’s modifier.

Those omissions defeat both motivation to combine and reasonable expectation of success.

Accordingly, the challenges to independent claims 1, 6, 11, and 16-18 fail on the merits.

1. Ground 1 Independently Fails Because Prabhavalkar is Not Analogous Art

Petitioner’s Ground 1 fails at the threshold because **Prabhavalkar** (Ex1005) is not analogous art. A reference is analogous only if it is from the same field of endeavor or is reasonably pertinent to the problem faced by the inventor. *In re Bigio*, 381 F.3d 1320, 1325 (Fed. Cir. 2004) (“Two separate tests define the scope of analogous prior art: (1) whether the art is from the same field of endeavor, regardless of the problem addressed and, (2) if the reference is not within the field of the inventor’s endeavor, whether the reference still is reasonably pertinent to the

particular problem with which the inventor is involved”). **Prabhavalkar** satisfies neither requirement.

Prabhavalkar is **not in the same field of endeavor** as the ’561 Patent. The ’561 Patent is directed to real-time accent correction and audio generation through an integrated pipeline that divides speech into input chunks, extracts acoustic features and reduced-accent linguistic features from the input chunk and context, receives a speaker embedding, and generates an output audio chunk based on those inputs. Ex1001, Abstract, Cl. 1. **Prabhavalkar**, by contrast, is an automatic speech recognition (“ASR”) reference that processes speech frames through a sequence-to-sequence encoder-attender-decoder architecture to generate recognition outputs, such as word elements or transcriptions. Ex1005, Abstract, 1:15–25, 2:41–3:56, 6:56–7:12, 11:19–57. Its chunking and context are used to improve streaming recognition, not to modify accents, extract reduced-accent linguistic features, receive speaker embeddings, synthesize accent-corrected audio, or generate output audio chunks.

Prabhavalkar is also **not reasonably pertinent to the problem addressed by the ’561 Patent**. The ’561 Patent uses chunks and context as part of a real-time accent-correction pipeline that extracts acoustic and linguistic information and uses those features, together with a speaker embedding, to generate accent-corrected

output audio. Ex1001, Abstract, Cl. 1. **Prabhavalkar** addresses a different problem: enabling an ASR model to process current and prior speech frames without waiting for a complete utterance and to output transcription-related recognition results. Ex1005, 6:56–7:12, 9:8–10:65, 11:19–57, 18:1–41. A PHOSITA seeking to build the claimed accent-correction pipeline would not have looked to **Prabhavalkar**’s ASR chunking and temporal context to solve the problem of extracting reduced-accent linguistic features for downstream audio generation.

Because **Prabhavalkar** is **not analogous art**, it cannot supply the claimed chunking or context limitations in Ground 1. Petitioner’s reliance on **Prabhavalkar** is hindsight-driven, and Ground 1 fails at the threshold.

2. Petitioner Fails to Show the Examiner-Identified Inventive Concept of the Spectrum Representation and the Spectrum Representation’s use in the Vocoder

A central problem with Ground 1 is that it attempts to reconstruct the very feature the Examiner recognized as missing from the prior art. During prosecution, the Examiner identified “*synthesizing a spectrum representation based on the acoustic features, the linguistic features, and a speaker embedding for a human speaker*” as not taught, disclosed, or suggested by the prior art of record in combination with the other limitations. Ex1002 at 176-77. The Petition illogically

conflates Feinauer's spectrogram analysis with [Fan](#)'s runtime melspectrogram synthesis, and as a result, the combination **would not result in the claimed invention**. Pet. at 20-27.

A PHOSITA would have understood that Fan's standard spectrogram uses a linear frequency scale with uniformly spaced frequency bands, while Feinauer's melspectrogram applies weighting that represents lower frequencies with greater detail and compresses higher frequencies. Therefore, the alleged architectural similarity between Feinauer and [Fan](#) in this respect is a mischaracterization.

In reality, Feinauer discloses spectrograms used for analyzing an incoming speaker's accent or dialect, **not for use in outputting audio during runtime**. Ex1004 ¶[0099]. [Fan](#)'s melspectrogram, by contrast, performs a different job as a required **intermediate representation** in [Fan](#)'s audio-generation pipeline; [Fan](#)'s neural network outputs a mel spectrum or mel feature, and [Fan](#)'s vocoder then converts that feature into a playable waveform. Ex1006 ¶¶[0097], [0099]. Thus, [Fan](#)'s vocoder is designed to operate on that specific kind of melspectrogram input. Feinauer's spectrogram has no runtime spectrum representation for the vocoder to operate on. Ex1004 ¶[0099]. There is **no rationale** for importing the vocoder from [Fan](#) into Feinauer's system. Feinauer **does not disclose** a pipeline in which the

system generates a new melspectrogram and then uses a vocoder to reconstruct audio from that melspectrogram.

Instead, Feinauer modifies live audio communications directly through its accent/dialect modification process. Ex1004 ¶¶[0007], [0036]. This modification of live video communications of Feinauer thus processes incoming audio, as opposed to [Fan](#), which teaches implementing a melspectrogram generated for output audio synthesis. As a result, the proffered motivation to combine is seriously flawed, as [Fan](#)'s melspectrogram is a building block for generating output audio, whereas Feinauer's spectrogram is merely a tool for analyzing input audio. [Fan](#)'s spectrogram/vocoder structure therefore does not naturally plug into Feinauer's architecture. Thus, even if a POSITA were to try and combine these elements, the results would not be the claimed invention—it would be chaos. [Fan](#)'s vocoder would have no operative input in Feinauer absent a wholesale redesign of Feinauer to generate the very spectrum-representation synthesis pathway claimed in the '561 Patent.

With respect to the petition's argument that “[a] PHOSITA would have understood *Feinauer*'s spectrogram of MFCC (mel frequency cepstral coefficients) teaches a melspectrogram”, a melspectrogram that is output by a neural network **is**

fundamentally different from a mel spectrogram that appears as an intermediate step in MFCC computation. Pet. at 32-33.

In [Fan](#), the mel spectrogram is an output of the neural network, passed to the vocoder as a key representation for audio synthesis. Ex1006 ¶¶[0097], [0099]. It plays a direct and essential role in runtime signal generation. *Id.* ¶[0099]. In contrast, the melspectrogram within an MFCC computation is merely a transient intermediate. It exists only momentarily as part of the transformation process and is immediately consumed: it is not preserved or used beyond enabling the next calculation. Ex1004 ¶[0099].

Accordingly, even if Feinauer can be said to “implicitly” compute a melspectrogram³, that computation occurs only as an internal and temporary byproduct that does not constitute a retained or functionally operative representation in runtime synthesis of the type required by [Fan](#)’s vocoder-based architecture.

³ The Petition argues “*Feinauer* teaches to ‘convert the audio sample waveforms into a spectrogram of MFCC[s] [mel-frequency cepstral coefficients].’ Pet. 32 (citing Ex1004, [0099], [0095]).

Because [Fan](#)'s vocoder is architecturally dependent on receiving a melspectrogram as input (Ex1006 ¶¶[0097], [0099]), it **cannot be isolated and integrated into Feinauer on its own**; a PHOSITA would have lacked both motivation and a reasonable expectation of success in attempting to incorporate [Fan](#)'s vocoder into Feinauer's system.

The **vocoder is needed** in [Fan](#) to perform an inverse transformation on the mel-spectrum to convert it back into a waveform (Ex1006 ¶[0099]), but Feinauer's **accent modifier operates directly on the incoming audio** waveform during runtime (Ex1004 ¶¶[0007], [0036]), so there is **no intermediate representation that requires transformation** and, accordingly, **nothing for a vocoder to invert** back into audio. **Thus, importing [Fan](#)'s vocoder would be non-operational without significant redesign, that would not advance the point of either [Fan](#) or Feinauer's disclosures.** It would require redesigning Feinauer to create a new intermediate melspectrogram output that Feinauer's system does not otherwise need.

Thus, [Fan](#) and Feinauer use spectrograms for **different and unrelated purposes which are not interchangeable**:

- [Fan](#) uses a melspectrogram as part of an audio-generation pipeline that requires a vocoder;

- Feinauer uses spectrogram-related information only to analyze or modify incoming speech.

Because Feinauer does not create a melspectrogram that must be converted back into audio, [Fan](#)'s vocoder has no natural role in Feinauer's system. Combining [Fan](#)'s spectrogram/vocoder architecture with Feinauer would require an unnecessary redesign, not a straightforward substitution. And even if a POSITA attempted to combine these elements, the results **would not be the claimed invention.**

3. Petitioner Fails to Show Extracting Linguistic Features with a Reduced Accent

Claim [1(a)] requires “*extracting, from a chunk of a stream of chunks of an input audio signal, ... linguistic features, wherein the linguistic features have a reduced accent.*” Equivalent language can be found in Claims [6(a)], [11(a)], [16(a)-(b)], [17(a)-(b)], and [18(a)-(b)]. That limitation is part of the claimed real-time accent-correction architecture, which requires chunked processing, context-based extraction of both acoustic and linguistic features, linguistic features with a reduced accent, receipt of a speaker embedding, and generation of an output audio chunk-based on that claimed combination of inputs. Ex1001 at Cl. 1. Petitioner **does not identify** any independent reference containing that architecture, nor does

Petitioner **provide a reasoned rationale** explaining **how** a PHOSITA would have combined the prior art to achieve it, let alone **why**.

Instead, Ground 1 pulls disconnected pieces from different systems:

- Feinauer for a general accent and dialect modification system;
- **Prabhavalkar** for chunking; and
- Fan for style/acoustic features, extraction, and embeddings.

Pet. at 18-20. **None** of these references teaches the independent claims, particularly extracting, from a chunk of a stream of chunks of an input audio signal, linguistic features, wherein the linguistic features have a reduced accent, and the Petition does not sufficiently explain **how the combination would achieve this result**.

- Petitioner does not show that Fan's PPG extractor could be applied to **Prabhavalkar**-style chunks and then inserted into Feinauer's accent/dialect modification framework.

The Petition offers an incomplete, irrational, and impermissible hindsight reconstruction. The Examiner already recognized that the prior art did not teach the claimed extraction of linguistic features with a reduced accent from both the input chunk and context. Petitioner's Ground 1 fails to show how the proffered prior art

overcomes this gap in the prior art, and thus **would not result in the claimed invention.**

4. Petitioner Fails to Show Speaker-Embedding Integration

Independent claims 1, 6, 11, and 16-18 require use of a speaker embedding for a human speaker. Ex1001 at Cls. 1, 6, 11, and 16-18. Petitioner relies on [Fan](#)'s style embedding vector. Pet. at 21-23. But [Fan](#)'s style embedding is **not** a universally compatible value that can be inserted into any accent-modification system. As the style embedding vector is not a pre-calculated value, capable of being freely extracted and implemented in other systems, the combination proposed by the Petition (Pet. at 20-23) would require incorporating the complete style separation model. However, the Petition provides no explanation as to **whether** Feinauer's system and [Fan](#)'s style separation model are compatible and if they are, **how** the incorporation would be effectuated, let alone how incorporating [Fan](#)'s style separation model into Feinauer's system would result in the claimed limitation.

The technical failure is not merely that Petitioner has mislabeled [Fan](#)'s style embedding as a speaker embedding. The failure is that the Petition never identifies any component in Feinauer that could use that embedding in the manner required by Claim [1(b)]. Claim 1 requires the speaker embedding to be one of the inputs

used to synthesize a spectrum representation, together with acoustic features and reduced-accent linguistic features. Feinauer does not disclose such a synthesis module, does not disclose combining a neural speaker/style embedding with linguistic features and acoustic features, and does not disclose generating a spectrum representation for downstream vocoder synthesis. Thus, even if [Fan](#)'s style embedding were imported, the proposed system would still lack the claimed architecture that gives that embedding functional significance and **would not result in the claimed invention.**

[Fan](#)'s style embedding vector is the output of a specific neural-network process. In the style conversion embodiment, the style separation model extracts timbre and sub-feature vectors from a third voice; user-adjusted weights may be applied to sub-features such as prosody, accent, and speaking speed; and a multihead attention structure outputs a style vector, for example a 256-dimensional style embedding vector. Ex1006 ¶¶[0087]–[0095]. That embedding is then used with a PPG feature in [Fan](#)'s voice fusion model. Ex1006 ¶¶[0008]–[0009], [0094]–[0095].

Feinauer is organized differently. Its accent and dialect modifier relies on word-level mapping and static pre-trained parameters. Ex1004 at ¶¶[0047]–[0057], [0061]–[0078], [0104]–[0117].

The Petition never explains **how** [Fan](#)'s embedding-producing neural network interfaces with Feinauer's modifier because there is no indication that Feinauer has the infrastructure to process high-dimensional continuous vectors. Incorporating [Fan](#)'s embeddings would require fundamental architectural components that are entirely absent from Feinauer's disclosure. This gap can be explained by the fact that it simply does not make sense to make such a combination.

In proposing the combination of Feinauer and [Fan](#), the Petition neglects to confirm whether Feinauer has the required hardware to run [Fan](#)'s software. Instead, the Petition simply acknowledges that “[a] PHOSITA would have recognized the modification constitutes applying prior art elements (*Feinauer*'s translation model and [Fan](#)'s style embedding vector) to a known data storage and retrieval methods with predictable results”. Pet. at 25. There is nothing included that remotely suggests that Feinauer's hardware is compatible with the [Fan](#)'s software. The fact that things are known is not enough to conclude they are interchangeable. *Virtek Vision Int'l ULC v. Assembly Guidance Sys., Inc.*, 97 F.4th 882, 887–88 (Fed. Cir. 2024) (“It does not suffice to meet the motivation to combine requirement to recognize that two alternative arrangements ... were both known in the art.”). The provided motivation to combine is lacking.

Hence, the Petitioner fails to establish the speaker-embedding limitation in the proposed combination. The Petition only explains that it would have been obvious to a PHOSITA to combine the “method of Feinauer ... [with a] speaker embedding (style embedding vector) to modify the accent (dialect) of the output audio signal.” Pet. at 24-25. There is no discussion of the means by which the model would be incorporated, or explanation of how this incorporation would actually result in the claimed invention. This is insufficient to establish an effective combination of references.

5. Petitioner Improperly Treats Fan’s Distinct Embodiments as Interchangeable

[Fan](#) describes distinct selectable modes with corresponding voice conversion networks: style conversion, dialect conversion, and voice enhancement. Ex1006 ¶¶[0005]–[0006], [0045], [0055]–[0060]. The Petition treats those embodiments as a single pool of interchangeable components. Pet. at 16-18, 20-25. [Fan](#) does not. Ex1006 at ¶¶[0009], [0018].

For Claims **[1(a)]**, **[6(a)]**, **[11(a)]**, **[16(b)]**, **[17(b)]**, and **[18(b)]**, the Petition relies upon [Fan](#)’s style separation model to meet the limitation requiring the presence of an “extracting ... acoustic features” and also uses the style separation model in cobbling together claim 1(b)’s “speaker embedding”. Pet. at 16-18.

However, the style separation model operates in completely different ways in Fan's style enhancement embodiment and dialect conversion embodiment. See Ex1006 at ¶¶[0009], [0018]. To overcome this hurdle, the Petition simply states, without further explanation, that "[a] PHOSITA would have recognized descriptions related to any hardware and/or software related to Fan's modules described functionality in Fan's style enhancement embodiment would have been equivalently applicable to the second dialect conversion embodiment as well." Pet. at 24. This is insufficient. There is **no** explanation as to **why** a PHOSITA at the time of the invention would have found this to be true.

In Fan's style conversion embodiment, the terminal obtains a first voice and a third voice, extracts a PPG feature from the first voice, extracts a style feature from the third voice using a style separation model, and inputs the style feature and PPG feature into a voice fusion model. Ex1006 at ¶¶[0008]–[0009], [0076]–[0094]. The style separation model extracts timbre and sub-features such as prosody, accent, and speaking speed, and the multihead attention structure outputs a style embedding vector, for example a 256-dimensional style embedding vector. Ex1006 at ¶¶[0087]–[0095].

In Fan's dialect conversion embodiment, the system uses a dialect conversion network. The first voice is a voice of a first dialect, and the second

voice is a voice of a second dialect. Ex1006 at ¶¶[0016]–[0019], [0068]–[0071]. Fan further discloses that the dialect conversion network may include multiple dialect conversion models and may preserve style by inputting a style feature of the first voice along with the feature information of the first voice. Ex1006 at ¶¶[0017]–[0019], [0127], [0162], [0166]–[0167].

Petitioner does not explain *why* a PHOSITA would take the style separation model and style embedding pathway from Fan’s style-conversion embodiment, combine it with Fan’s dialect-conversion pathway, and then graft the result onto Feinauer. Those operations are **materially different**:

- one extracts a target style from a third voice;
- the other converts dialect and may preserve source style.

Petitioner proffers no evidence that the style separation model would be able to function in its acoustic extraction role within the dialect conversion embodiment architecture, as would be necessary to meet the limitation of the ’561 Patent. The Petition’s assertion without any evidentiary support that the two embodiments are “equivalently applicable” is both misleading and inaccurate. Pet. at 24. They are clearly different, and more importantly not clearly equivalent. It would not have been obvious to a PHOSITA at the time of the invention to combine the two

embodiments—the hindsight reconstruction **would not result in the claimed invention**. Petitioner’s unsupported cross-embodiment reconstruction renders Ground 1 fatally flawed.

6. Petitioner Fails to Show Context-Based Feature Extraction

Independent claims 16-18 require two context-based extraction steps. The acoustic features extraction process must extract acoustic features “from the input chunk and a context associated with the input chunk.” Ex1001 at Cl. 16-18. The linguistic features extraction module must also extract linguistic features with a reduced accent “from the input chunk and a context associated with the input chunk.” *Id.* Petitioner cannot satisfy those limitations through manipulation of chunking in **Prabhavalkar** and feature extraction in [Fan](#).

As discussed above, **Prabhavalkar**’s context is ASR context. Ex1005 at 9:20–36. Its encoder generates encoded feature representations; its attender generates a context vector by attending to encoder outputs; and its decoder uses that context vector to output probability distributions over word elements for transcription. Ex1005 at 9:8–10:65, 11:19–57. In streaming implementations, **Prabhavalkar**’s attender may compute attention based on speech frames in the current chunk and one or more previous chunks, including at least ten previous

chunks. *Id.* at 6:56–7:12. That context supports recognition, not accent-corrected audio generation.

Fan's feature extraction operates differently. Fan's voice feature extraction model extracts PPG features from a first voice. Ex1006 at ¶¶[0007], [0067], [0084]–[0086]. But Fan does not disclose extracting PPGs or style/acoustic features from an input chunk context. The attention mechanism disclosed in Fan, unlike that of **Prabhavalkar**, operates without a temporal dimension. *Id.* at ¶[0094]. Fan's system focuses on weighting an extracted set of feature vectors (*e.g.*, timbre vector, prosody vector, accent vector) rather than analyzing a temporal sequence of audio chunks. *Id.* at ¶[0090].

Petitioner's reliance on "attention" does not bridge the gap. Pet. at 38-40. **Prabhavalkar**'s attention is temporal ASR attention. Ex1005 at 7:5-16. It accepts as input speech frames in the current chunk or previous chunks, and outputs at least one probability distribution of word elements for each chunk. *Id.* at 6:56–7:12. Fan's multihead attention accepts as input a feature vector (*e.g.* timbre feature vector) and weighted sub-feature vectors and aligns them to output a style embedding vector. Ex1006 at ¶[0094]. These mechanisms serve different purposes, use different inputs, and generate different outputs. *See* Ex1005 at 2:44-47, 6:56-7:16; Ex1006 at ¶¶[0012], [0094]. The Petition never explains how

Prabhavalkar's temporal ASR lookback window would be inserted into **Fan**'s PPG extraction or style separation models. And unlike **Prabhavalkar**, **Fan** does not require or utilize a lookback structure; instead, it extracts all necessary information from a single voice sample in a unified operation. *Id.* at ¶¶[0005]–[0008], [0066]–[0071]; *see also id.* at ¶¶[0016], [0020], [0084], [0087]. As **Fan**'s attention mechanism acts upon an already computed set of features, it functions independent of the prior audio history.

Prabhavalkar's context has no identified destination in **Fan**. **Prabhavalkar**'s prior-chunk context is consumed by an ASR attention mechanism that attends over encoder outputs to assist recognition. **Fan**'s PPG extractor and style-separation models, however, are not disclosed as receiving a current audio chunk plus cached prior-chunk context, nor are they disclosed as using temporal ASR attention over prior chunks. **Fan**'s multihead attention instead operates on already-extracted style vectors and weighted sub-feature vectors. Thus, the Petition never identifies a feature-extraction module that accepts the claimed input chunk and associated context and outputs acoustic features or reduced-accent linguistic features, **failing to result in the claimed invention.**

Given these fundamental differences, the Petition fails to provide a persuasive rationale for why a PHOSITA would seek to integrate **Prabhavalkar**'s

temporal chunking into [Fan](#)'s weighting architecture. *See* Pet. at 38-40. The two systems employ “attention” in fundamentally different capacities: [Prabhavalkar](#) uses it to navigate an evolving temporal sequence, whereas [Fan](#) uses it to weight static acoustic embeddings. Because these mechanisms serve distinct functional roles and operate on different types of data, the mere commonality of the term “attention” is insufficient to motivate the combination of references.

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

Ground 1 also challenges dependent Claims 2, 7, and 12. Those dependent-claim offerings do not cure the deficiencies in Petitioner's showing for independent Claims 1, 6, and 11, respectively.

Claims 2, 7, and 12 depend from Claims 1, 6, and 11. Even if Petitioner could identify isolated teachings for a melspectrogram and acoustic features, those teachings do not establish the required base pipeline of Claim 1. Petitioner must first show the claimed chunk-based extraction of acoustic and linguistic features, speaker-embedding integration, linguistic features with a reduced accent, and output-chunk generation by a vocoder. It has not done so.

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

B. Grounds 2-4 Fail on the Merits

1. Grounds 2-4 Fail Because They Rely on Ground 1

Grounds 2-4 fail at least because they depend on Ground 1's defective base combination. Ground 2 adds Sargsyan for Claims 3, 8, and 13; Ground 3 adds Kalinli for Claims 4, 9, and 14; and Ground 4 adds Watts for Claims 5, 10, and 15. Pet. at 43-54.

Because Petitioner has not shown that Ground 1 satisfies independent claims 1, 6, 11, and 16–18, Petitioner cannot show that the dependent claims are unpatentable in Grounds 2-4. The additional references are used only for additional dependent limitations. They do not cure Ground 1's failures regarding chunking, extraction, and speaker-embedding integration. The Board can deny Grounds 2-4 because they depend on Ground 1.

2. Ground 4 Also Fails Because Watts Does Not Teach the Claimed Time-Shift Parameter

Ground 4 challenges Claims 5, 10, and 15 and adds Watts for computational graph/time-shift limitations. Watts discloses a noise-robust speech recognizer using high-resolution cochlear modeling, stream separation, and phoneme classification. Ex1008, Abstract, 1:16–17, 3:11-15, 4:37–47, 7:5–20. However, Watts does not disclose a “time-shift parameter” indicating a difference in timestamps of a chunk and input or output data of a module, nor does it assign acoustic features a lower

time-shift parameter than linguistic features **Ground 4 does not offer any fix for the failures of Ground 1 discussed above..**

In any event, the Petition relies upon Watts' 1 ms and 150 ms disclosures to meet this limitation (Pet. at 51-52), but these are not time-shift parameters. In Watts, the 1 ms value refers to the update rate/time-scale of the spectral envelope feature extraction: Watts states that the recognizer extracts spectral envelope "on a much finer time-scale (e.g., 1 ms)" and identifies a "600-tap Spectral Envelope updated every 1 ms." Ex1008, 3:11-28, 4:43-47, 6:63-7:5. Furthermore, the 150 ms value refers to the fixed context window supplied to the phoneme classifier for spectral envelope, pitch/voicing, and transient-detection inputs. Ex1008, 7:11-15. These values describe the temporal scale or context used by Watts' feature extraction and classification process; they do not represent a calculated difference between a specific input-audio chunk's timestamp and associated module input or output data.

Moreover, Watts does not disclose providing any module with a time-shift parameter indicating a timestamp difference, nor does Watts disclose assigning different time-shift parameters to acoustic features versus linguistic features. Ground 4 therefore fails.

3. Ground 3 Also Fails Because Kalinli's Parallel Acoustic Filters Do Not Teach the Claimed Parallel Processing

Ground 3 challenges Claims 4, 9, and 14 and adds Kalinli for parallel processing. Kalinli teaches tone/intonation recognition so that multiple processors may extract different acoustic features from an input sound window in parallel. Ex1009, Abstract, ¶¶[0017], [0021]–[0023], [0048]–[0051]. Petitioner's motivation to combine Kalinli is critically unsupported. **Ground 3 does not offer any fix for the failures of Ground 1 discussed above.**

In any event, Kalinli discloses a method for tone/intonation recognition. It does so by analyzing the sound signal for auditory patterns, summarizing those patterns, and then using a machine learning algorithm to classify tonal characteristics. Ex1009 ¶¶[0017]–[0034]. Thus, it is a recognition/classification system, not a speech-generation or accent-correction system: Kalinli does not synthesize speech, generate a spectrum representation, use a speaker embedding, or produce output audio chunks.

Petitioner assumes that Kalinli's parallel-processing disclosure could be lifted from its tone/intonation recognition system and inserted into the alleged Feinauer–[Prabhavalkar](#)–[Fan](#) pipeline through mere “software reprogramming”. Pet. at 49. That assumption is unsupported. Kalinli's parallel processing is designed around its own feature extraction modules; the parallelization Kalinli describes is

not a generic programming pattern but is tightly coupled to its specific structure and purpose. *See e.g.* Fig. 4, [0049]-[0051]. Kalinli's parallel architecture is directed to analyzing the same sound signal using several related audio measures, such as intensity, frequency contrast, temporal contrast, and orientation. Those operations are all directed to extracting acoustic features for recognition purposes: they do not extract linguistic features, and they do not support speech synthesis or vocoder-based audio generation. Ex1009, ¶¶[0021]–[0026], [0050].

That is materially different from Petitioner's proposed use of Kalinli. Petitioner invokes Kalinli to justify parallel processing in the alleged Feinauer–Prabhavalkar–Fan accent correction pipeline, particularly for parallelized extraction of acoustic and linguistic features. Pet. at 47-48. But Kalinli does not teach parallel extraction of distinct acoustic and linguistic feature streams, nor does it explain how such streams would be coordinated with later synthesis and vocoder steps. Ex1009, ¶¶[0021]–[0026], [0050]. Petitioner never explains how Kalinli's recognition-oriented parallel feature extraction would be integrated into the alleged pipeline, what modules would be parallelized, or what data would be exchanged between them. Pet. at 48-49.

Thus, Petitioner has not provided the required motivation to combine or reasonable expectation of success. Ground 3 therefore fails.

4. Ground 2 Also Fails Because Sargsyan Does Not Cure the Missing Claimed Pipeline

Ground 2 challenges Claims 3, 8, and 13 and adds Sargsyan for different input and output sample rates. Pet. at 43-47. The Petition does not explain how Sargsyan's bandwidth-expansion process would be integrated into the alleged Feinauer–[Prabhavalkar](#)–[Fan](#) system. *Id.* **Ground 2 does not offer any fix for the failures of Ground 1 discussed above.**

In any event, it fails to identify which signal in the Feinauer system processing pipeline serves as the “input” at the first sample rate and which serves as the “output” at the second sample rate. Pet. at 43-47. Further, there is no discussion regarding whether the 8kHz to 16kHz framework that is taught in Sargsyan would be applicable to the signal in the Feinauer system. *Id.* In fact, there is no mention of sample frequency at all in Feinauer, [Prabhavalkar](#), or [Fan](#). Ex1004, Ex1005, Ex1006.

Moreover, Sargsyan's teachings on sample-rate conversion are directed to minimizing bandwidth degradation in cellular networks, not to accent correction. Ex1007, Abstract, [0027]. The Petition does not explain why Sargsyan's cellular-bandwidth enhancement framework would be applied to accent correction, especially where neither Feinauer, [Prabhavalkar](#), nor [Fan](#) supply the missing sample-rate architecture in the proposed combination. Ground 2 therefore fails.

VI. CONCLUSION

Ground 1 is the foundation of the Petition, and Petitioner does not show a supported motivation to combine Feinauer, [Prabhavalkar](#), and [Fan](#) or a reasonable expectation of success. Because Grounds 2-4 depend on Ground 1's defective base combination, those grounds fail as well. Furthermore, Grounds 2-4 each also fail for defective combinations.

For the foregoing reasons, Patent Owner respectfully requests that the Board deny institution of post-grant review of Claims 1–18 of the '561 Patent.

Respectfully submitted,

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

Dated: June 25, 2026
New York, New York

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

CERTIFICATE OF WORD COUNT

The undersigned certifies that the foregoing PATENT OWNER'S PRELIMINARY RESPONSE complies with the applicable type-volume limitation in 37 C.F.R. § 42.24. According to the word-processing system's word count, the brief contains approximately 10,539 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 25, 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 25, 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. Bruesh (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 25, 2026
New York, New York

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