Episode Book

EPISODE #6623

Teaching Five AI Voices Israeli Hebrew

PUBLISHED	16 November 2026 at 00:37 UTC
DURATION	28:53
LISTEN	myweirdprompts.com/episode/multi-voice-hebrish-tts-48049/
HOSTS	Corn, Daniel, Herman, Hilbert
SCRIPT MODEL	anthropic/claude-sonnet-5.5
VOICES	omnivoice
PIPELINE	V5.33
BOOK GENERATED	16 November 2026 at 02:53 UTC

SYNOPSIS

A speech model can nail an English sentence and still butcher a single Hebrew word in the middle of it. Meorav Multi Voice solves that join by training Qwen3-TTS on 461 code-switched takes where only the Hebrew spans are graded, keeping the pronunciation without leaking the speaker's voice. Five synthetic accents from OmniVoice—American, British, and Australian—share that Hebrew training for about $4.40, then star in a ten-minute radio comedy about Israeli bureaucracy.

TRANSCRIPT



Corn

What does a speech model do when an English sentence has a Hebrew word in the middle of it, and it has never heard anyone say that word properly? It guesses. Usually it guesses like a tourist reading a menu. That's the itch behind today's episode, and Daniel is here in the studio to scratch it.



Daniel

This is the second side adventure in a few days, separate from the work on your own voices. I've been building something I call Hebrish. It's a text-to-speech model that speaks English, but when a Hebrew word turns up, it says it the Israeli way. The project is called Meorav, which means mixed, and the new version is Meorav Multi Voice. Get into how it works, and then I'll play you something I made with it.



Herman

Mixed is right, and it's a hard problem for a reason people don't expect. The model isn't failing at Hebrew. It's failing at the join. Say a sentence like "I need to stand in line at the Misrad HaPnim." Every other word is perfectly ordinary English, and then three syllables arrive that the model has only ever seen spelled in Latin letters, with no clue how an Israeli would stress them or where the throat goes.



Corn

Daniel, you started with your own voice, though. Meorav was already speaking Hebrish. Why go looking for other voices at all?



Daniel

Honestly, a few reasons, and they're not very noble. The earlier versions were fine-tunes in my voice, and that version is now split out as Meorav Daniel. But when I tried a designed American voice, the bland one, I found it slightly grating. I'm sorry, but I did. Then I tried Irish, and that was too distinctive. It pulled the ear away from the Hebrew. What I wanted was something more fun, and I wanted a female voice, because nothing I had was one.



Herman

That's a real design constraint, actually. If the accent is too strong, the listener hears the accent first and the pronunciation second. You want voices that are distinctive enough to be characters, but not so loud that they drown out the thing you're trying to show off.



Corn

Irish too distinctive, says the Irishman. Noted. So you built five of them. Where did the voices come from?



Daniel

They're designed. Not recorded, not cloned from anyone. I used a voice design model called OmniVoice and fed it text prompts, something like "female, young adult, moderate pitch, American accent". I swept four accents, a couple of seeds each, and picked five. An American male from an earlier pilot, called A three, then an American woman, a British man, a British woman and an Australian man. None of them is a real person.



Herman

And OmniVoice came into the story earlier, didn't it? As the thing that failed.



Daniel

It did. On the fourteenth I tried to teach OmniVoice itself to say Hebrew inside English, and it couldn't. I gave it Hebrew with the vowel marks, and it got one word out of eleven right. One. Then I tried spelling the sounds out with the phoneme overrides, the ARPAbet route, and that got about seventy-three percent, which sounds respectable until you listen. It never made the guttural sound, and the word came out disjointed, like it had been taped together.



Corn

Seventy-three percent of the sounds, and zero percent of the soul.



Herman

The guttural is the whole signature, though. If the ch in chag comes out as a k, you've said a different word, as far as a native ear is concerned.



Daniel

That's the problem. So OmniVoice stayed in the job it's good at, which is inventing voices from a description. The actual speaking model is Qwen three TTS, the one point seven billion parameter base. I took my Meorav Daniel weights, put the original speaker encoder back, and that meant a new voice is just a short reference clip. The model hears the clip, makes a voiceprint, and speaks in it.



Herman

Which raises a problem I'd like to hear you solve out loud. You want the Hebrew to come from you, because you have the pronunciation. But you don't want the voices to sound like you.



Daniel

That was the clever bit, and I'll admit it wasn't my idea in the first draft. I have four hundred and sixty-one takes where I code-switch, English sentences with Hebrew words in them. In training, I only let the model learn from the spans where the Hebrew words are. The English around them is masked out of the loss. So it's graded on how I say the Hebrew and not graded on anything else I do.



Corn

So it learns the throat, not the person.



Herman

The mask matters because a speech model soaks up everything. Rhythm, breath, the way you land a sentence, the Irish in your vowels. Grade it on whole sentences and your voice leaks into all five characters, and you get five Daniels in different hats. Grade it only on the Hebrew spans, and the one thing it can take from your takes is how the Hebrew word is made.



Daniel

The marking was its own job. Word labels came from Sonnet subagents, timings from a forced aligner, and I'd cleaned the takes up with a speech enhancement model first. And the English comes from the designed voices. Each of the four new ones got two hundred and forty-five English lines, generated by OmniVoice, about half an hour each, with four small GPUs doing it in parallel. Those lines are trained in full. Altogether it's two thousand one hundred and forty-seven rows, and every row is conditioned on its own speaker's voiceprint.



Herman

Can I ask the obvious one? Why one model for five voices? You could have trained five.



Daniel

I asked that myself, mid-run. I asked whether doing them together was cheaper, and it was. The Hebrew takes are the expensive part, and they get learned once and shared. And having several distinct voices gives the model a reason to listen to the voiceprint instead of just collapsing into one. All in, the cost was about four dollars forty, an estimate from GPU time, not something I've checked against the bill.



Corn

Four forty. That's less than the coffee I'd need to stay awake while it trained. Which, to be fair, I'd have slept through.



Herman

Putting them together is where the saving lives. The preparation of the Hebrew data is the part you pay for once, and each extra voice only costs its own English lines. Corn, you'd have slept through the coffee too. The recipe was modest, wasn't it? Three epochs, a small learning rate, one big GPU.



Daniel

One A one hundred and about an evening. It finished around half past one in the morning on the sixteenth. After each epoch it rendered ten test lines in each voice, so a hundred and fifty clips, and I listened. I'll tell you what I said out loud at the time. "They sound great."



Herman

The final loss was around one point two to one point three by the last epoch. That tells me it trained. It says nothing about whether anyone would believe the result, which is why the listening matters.



Corn

And then, being you, you didn't stop at test lines.



Daniel

No. I wanted to know if it could carry a longer piece. So I wrote a radio comedy, about ten minutes. It's called "The Teudat Zehut: a comedy in four queues," and it's about a new oleh's day of Israeli bureaucracy. A teudat zehut is the ID card, the thing you need to do anything else. I narrate it in my own Meorav voice. Bubbe Rivka is an American grandmother, obsessed with lokshen and chicken soup, whom everyone says sounds twenty-five. Nigel is the Misrad HaPnim clerk, and he's from Manchester. Fiona is at the bank. Bazza is from Perth. And the American male voice, A three, does the phone menus and the ticket machine.



Herman

A clerk from Manchester at the Interior Ministry is already a better premise than most sitcoms. One thing to listen for, before we play it. The Hebrew words should sound the same in every mouth, and only the voice around them should change. That's the claim the whole build rests on.



Corn

I'd watch a sitcom where the machine is the only one who speaks clearly. Five parts, I'm told. Daniel, you're sitting through it with us?



Daniel

I'll be sitting here pretending I've never heard it. The day starts at the ministry, in a queue that doesn't move.



Herman

Then the ticket machine can have the floor. Let's hear the first queue.



Corn

Well, I was ready to be polite about that and now I'm just laughing at a soup delivery.



Herman

Bubbe Rivka brought the lokshen into a government building like it was a hostage exchange. And she really does sound twenty-five. The line about the age is true to the ear. It's the voice of someone who just graduated, saying things like "darling, sit".



Daniel

That's the designed voice. I asked for young adult and a certain energy, and she turned into someone's grandmother. I didn't plan the contrast.



Corn

Bazza being told "acharei hachagim" three times with that Perth lift on it. After the holidays. After the holidays. The man has been told that for what feels like a decade.



Herman

What I noticed is that the Hebrew didn't change with the voice. Rivka's acharei hachagim and Bazza's are the same Hebrew in two throats. That's the span masking doing exactly what it's supposed to. Daniel, were you surprised?



Daniel

A bit. Bazza especially. His accent is as Australian as it gets for everything else, and then the Hebrew word comes in and it sounds like it was picked up on a Jerusalem bus.



Corn

Whoever wrote Nigel knows a clerk. Nigel next, then.



Daniel

He's at counter three, and he's the one person in the building who can actually help. Over to Nigel at the counter.



Corn

Well. "Chagim all the way down." I'm going to be hearing that in my sleep, and I'm already asleep for most of the day.



Herman

The loop is elegant, though. The clerk needs a tor, the bank needs the teudat zehut, the teudat zehut needs the bank statement, and nobody is lying. Everyone is right, which is the cruel part of it.



Daniel

That's the scene I wrote first. It's the one I'd say is the least exaggerated.



Corn

Nigel is perfect, mostly because he sounds sorry. A Mancunian clerk apologising in Hebrew for a queue number. The sorry is doing a lot of work.



Herman

Corn, did you notice the tor? He says it the Israeli way, and then English around it. Tor. Not "tore". Your own ear would have been all over that, as the one who has lived in this town the longest.



Corn

I've lived here on a diet of leaves and hope. But yes, the tor was clean.



Daniel

That's the real test, honestly. If a Hebrew word is one syllable, you hear immediately whether the model understood or not.



Herman

The bank is next, and there you've got two new voices in one room. Fiona is British and Rivka is American, so you'll hear the same Hebrew pass between them.



Daniel

Fiona runs the counter, and Rivka turns up with a pot. Off to the bank we go.



Corn

The soup bribes the manager. I'd like to think there's a precedent somewhere, a small bank in Haifa, a bowl of chicken broth and a mortgage.



Herman

Fiona calling it the loop was my favourite thing in the scene. She says it the way a nurse says "the usual". Flat, tired, a bit pleased with herself.



Corn

And Rivka was at Woodstock with a kugel. That sentence is going on my gravestone, whenever the sloth in me gets around to dying.



Daniel

I wrote that line at one in the morning and I was not sure it was funny. Hearing it in her voice, it is.



Herman

It's funnier because the voice is so bright. Rivka's confidence carries things the writing alone wouldn't. That's a bit unnerving, if you think about it, as a thing we're being delivered by.



Corn

Herman, we are a donkey and a sloth with voices. We have no leg to stand on.



Daniel

Now it's Kupat Cholim, the health fund. You're on hold, A three is the phone menu, and the waiting music is a song I won't spoil. Cue the hedgehog hold music.



Corn

I was prepared for that to be a bit and it turned into a full minute of real hold music, with real despair.



Herman

The menu was the one that got me. A three reading "for Russian, press three" and then every option leads to the same place. That's a piece of design, not just comedy. You can hear how a neutral American voice makes the machine more machine-like.



Daniel

He's the neutral one. Nothing for the ear to hold onto, which is exactly what a phone menu wants.



Corn

A voice with no personality finally gets a job where that's the requirement.



Herman

Listening to these, I keep noticing they're voices that aren't us, doing a bit. Corn, do you hear them as colleagues or competition?



Corn

Colleagues who've never had to nap. I'm fine with the competition.



Herman

Then the offices closed, acharei hachagim, again. They kept coming back to it. It's a running gag, and it works because it's also true.



Daniel

Last scene. We're back at the Misrad HaPnim, with every loose thread in one room. And now the grand finale.



Corn

Rivka's look. That's the whole resolution, a look, and a stamp. I respect a story where the hero is a grandmother's glare.



Herman

It lands because the setup is airtight. She has been told no by every voice for ten minutes, and the stamp comes from nobody doing anything differently. The only thing that changes is her eyebrow.



Daniel

And then you have me narrating that I officially exist, and the actual card comes in the post, acharei hachagim.



Corn

Of course it does. You exist, but the proof is in the post, after the holidays.



Herman

I laughed out loud. Daniel, how much of that was cheating? Did you fix the pronunciations by hand before rendering?



Daniel

No. Same pipeline for everyone, and it's a first render, so I won't pretend it's perfect.



Corn

Good, because I noticed. The throat sound. In chag, or in achi, it wasn't always there. In a couple of places it was a soft h, a breath instead of a rasp.



Daniel

Yes, that's the known limitation. The guttural, the ch in chag and in achi, isn't always produced. It's better than where OmniVoice ended up, but I won't claim it's solved.



Herman

My guess about why, and it is only a guess: English gives the model hundreds of thousands of examples of h and k, and it has a few hundred takes of the real rasp. When it's unsure, it falls back on the nearest English sound.



Corn

A breath is the safe bet. Nobody gets sent home for a breath.



Herman

And you haven't benchmarked it. There's no metric behind "they sound great".



Daniel

Correct. It's me, listening, on one evening. No score. I could build something to measure it, but at the moment it's my ears, and I'm a biased judge because I wrote the jokes.



Corn

Reasonable. A comedy is a poor benchmark, anyway. Everybody laughs at their own Woodstock kugel.



Herman

So where does someone get it?



Daniel

The model is public on Hugging Face, under the Hebrish organisation, as Meorav Multi Voice. It's version one point oh. The licence is Creative Commons attribution non-commercial, and it's gated with automatic approval, so you click through and you're in. The voice reference clips are in the repository. And there's free demo audio in a Hebrish demos dataset, which is ungated, and this drama is in it. The Hebrish site has a page for the model and a blog post on the build.



Herman

Non-commercial is worth hearing twice. Play with it, build demos, make comedies, but the licence stops short of selling anything made with it. And since the voices were designed rather than cloned, there's nobody's real voice in the package.



Corn

And the first upload crashed, from what I read in your notes.



Daniel

It did. The script read the model card inside the cloud container, where the card wasn't. I fixed it by reading it locally, then checked every file's hash on the Hub against the original, and they matched.



Hilbert

Grating is the wrong word. Nobody can grate a voice. A voice is thin or it's thick. I did a client's answering machine last week, a Mr. Ostrowski, he runs a boiler repair outfit in Wolverhampton, and he wanted the greeting to say the name of his street, which is Snape Road, and every program I tried made it rhyme with tape. I said no. He wanted me to spell it phonetically in the text box, and I said, Mr. Ostrowski, I'm not spelling your own street wrong in front of your own customers. So I recorded it myself, Herman. Forty-one seconds, with a stop for the kettle, and the microphone doesn't sit right on a snout, so I did it standing. Stan Pruett could hold a dial tone for fifty-one seconds without a breath, and he's never let anyone forget it, and my greeting is still shorter than that. Mr. Ostrowski paid sixty pounds.



Corn

Sixty pounds for the street name done properly, and nobody had to learn the vowel in a silent letter. That's the going rate, then, for avoiding a rhyme with tape.



Herman

Snape Road is a nasty test word, mind. A silent letter, one syllable, and it's the one word on the recording he can't afford to get wrong.



Hilbert

The kettle's audible in the first take. Mr. Ostrowski rang to say he preferred that one. I charged him the same.



Herman

Question for you, Corn. If you were choosing the voice, who'd it be? Rivka, with the cheerful menace, or the clerk who apologises?



Corn

Nigel. I'd want the apologising. A voice that says sorry about the queue number is a voice I'd trust with my passport.



Daniel

I'd go Bazza, I think. Because he sounds like he's having a nice time, which I can't say for most of the people in the Misrad HaPnim.



Corn

There's a misconception I'd like to kill, which is that this is a Hebrew model. People will hear "Hebrish" and assume you can feed it a Hebrew paragraph.



Herman

It's an English model, and the Hebrew words have to sit inside English sentences. Whole Hebrew sentences aren't what it was taught. It learned the words you'd drop into English, the way an oleh does in the queue.



Daniel

Code-switching, not translation. It's how my own mouth works after all these years, and now there's a model that does it too.



Herman

That's the useful part for anyone building speech tools. A great deal of real speech, in Israel and in plenty of other places, has this shape: one language carrying a handful of words from another, and the words from the other language are exactly the ones a standard model flubs. The fix doesn't need a vast corpus. It needs a few hundred careful takes from one person who says the words right.



Corn

A few hundred takes and about four dollars. I've paid more for a bad sandwich in Tel Aviv.



Herman

Corn, do you think the five voices are the point, or is the point that you can swap the voice later and keep the Hebrew?



Corn

The second. Voices are costumes. The pronunciation is the actor, and it's the same actor in every costume. Today's other episode is about our own cast voices moving to adapters. That's us. This one is voices that aren't us, passing through a queue.



Herman

One thing I'm taking from the drama and the limits, and from the boiler man, is that the best test of a voice is a word it might get wrong. Short words, single syllables. If the model gets a plain tor right, the rest is a matter of time.



Daniel

Throat sounds are version two.



Corn

Thank you to Hilbert Flumingtop, who produces the show. This has been My Weird Prompts. If you want more, the website is my weird prompts dot com.



Herman

Daniel, thanks for bringing the queue in with you. I'm only sorry we couldn't get a stamp.



Daniel

Thanks, both of you. See you tomorrow.