Speech-to-Speech AI: The Voice Revolution

Speech-to-Speech AI: The Voice Revolution Transforming Australian Business in 2026

Category: Technology | **Read Time:** 12 min | **Date:** February 2026

TL;DR

Speech-to-speech AI models have matured dramatically in 2026. Unlike traditional systems that chain together speech-to-text and text-to-speech (creating latency and losing nuance), modern speech-to-speech models process and generate audio directly through a single model. This means near-instant responses, natural conversation flow, and the ability to detect emotion and intent from the speaker's voice. For Australian businesses, this unlocks powerful applications: 24/7 voice customer service, automated appointment booking, mass outreach campaigns, and voice assistants that actually sound human. Agentive partners with Heya, Australia's leading AI voice agent platform, to help organisations deploy voice AI that works.

Introduction: Voice AI Has Finally Grown Up

If you have ever called a customer service line and heard a robotic voice stumble through a menu, you know why most people despise automated phone systems. They are slow, frustrating, and about as natural as a conversation with a malfunctioning toaster.

That era is ending.

Speech-to-speech AI in 2026 represents one of the most significant leaps in how machines communicate with humans. We are not talking about incremental improvements to text-to-speech. We are talking about AI that listens, understands context and emotion, thinks, and responds naturally; all in real time, with latency measured in milliseconds.

At Agentive, we have been tracking this technology closely because it changes what is possible for Australian businesses. Combined with our partner Heya, Australia's smartest AI voice agent platform, we can now deploy voice AI solutions that genuinely transform how organisations interact with their customers.

This is not hype. This is production-ready technology that businesses are deploying today.

What Is Speech-to-Speech AI? (And Why It Matters)

Traditional voice AI systems work like this: your voice gets converted to text (speech-to-text), the text gets processed by an AI model, and then the response gets converted back to audio (text-to-speech). This three-step pipeline creates problems.

The Old Way:

- Latency builds up across multiple model calls
- Nuance and emotion in your voice are lost in transcription
- Responses sound robotic because text-to-speech has no context about the conversation
- Interruptions and natural turn-taking are handled poorly

The New Way (Speech-to-Speech):

- A single multimodal model processes audio input directly
- The model "hears" emotion, intent, and tone; not just words

- Responses are generated as natural speech with appropriate expression
- Latency drops dramatically (under 150 milliseconds in many cases)
- Natural conversation flow with proper turn-taking and interruption handling

OpenAI describes their approach: “The model thinks and responds in speech. It doesn’t rely on a transcript of the user’s input. It hears emotion and intent, filters out noise, and responds directly in speech.”

This is a fundamental architectural shift, not just a feature update.

The Technology Landscape in 2026

Several major players have advanced speech-to-speech capabilities significantly:

OpenAI’s GPT-Realtime

OpenAI’s Realtime API is now out of beta and production-ready. Their latest model, gpt-realtime, is specifically trained for voice agent applications.

Key capabilities: - Direct speech-to-speech processing through a single model - 18.6 percentage point improvement in instruction-following accuracy - 12.9 percentage point improvement in tool-calling accuracy - Support for remote MCP servers, image inputs, and SIP phone calling - Fine-grained voice customisation (“speak quickly and professionally” or “speak empathetically in a French accent”)

The mini model variant (gpt-realtime-mini) is optimised for low-latency real-time interactions, making it suitable for production voice agents that need to respond instantly.

ElevenLabs Conversational AI 2.0

ElevenLabs launched Conversational AI 2.0 in January 2026 with enterprise-focused features:

Natural turn-taking: Custom models designed to understand conversation flow, eliminating awkward pauses and unnatural interruptions.

Integrated RAG: Retrieval-Augmented Generation built directly into the voice agent architecture. Your agent can access your knowledge base with minimal latency; perfect for support agents accessing product information or medical assistants retrieving treatment guidelines.

Multimodality: Agents can communicate over text, voice, or both simultaneously. Define once, deploy everywhere.

Batch calling: Initiate multiple outbound calls simultaneously for alerts, surveys, or personalised messages at scale.

ElevenLabs reports their own documentation voice agent successfully handles over 80% of user inquiries across 200 calls per day.

Open-Source Progress

The open-source community has made remarkable advances:

- **Fish Speech V1.5:** Top choice for multilingual applications requiring highest quality
- **CosyVoice2-0.5B:** Real-time streaming with 150ms latency
- **Kokoro:** Just 82 million parameters delivering near-commercial quality; ideal for edge deployment
- **NeuTTS Air:** The world's first on-device, super-realistic TTS with instant voice cloning, running on laptops, phones, and even Raspberry Pis

For speech recognition, NVIDIA's Canary Qwen 2.5B tops the Hugging Face Open ASR Leaderboard with 5.63% word error rate, while Deepgram Nova-3 excels at real-time multilingual transcription in noisy environments.

Heya: Australia's AI Voice Agent Pioneer

While global players build general-purpose platforms, Australian company Heya has focused specifically on the needs of Australian small and medium businesses.

What makes Heya different:

Australian voice and understanding: Heya speaks with a natural Australian accent and understands local speech patterns, slang, and context. This is not a US platform with an Australian flag slapped on it.

Small business focus: While others target enterprise, Heya is designed for local cafes, tradies, medical practices, and shop owners who keep communities thriving.

24/7 availability: Heya handles every call and message, turning after-hours inquiries into business opportunities. While you rest, your business stays open.

Practical features: - Handles common questions instantly with zero wait times - Books appointments based on real-time availability and syncs to your calendar - Runs mass outreach campaigns that would take days in just minutes - Detects customer emotional tone (positive, negative, neutral) and adjusts responses accordingly

Real results: Heya claims to save businesses 20+ hours per week and handle 80% more customer interactions.

Agentive partners with Heya because their platform aligns with our philosophy: enterprise-grade AI made accessible for Australian businesses of all sizes.

Real-World Applications for Australian Businesses

Speech-to-speech AI is not a solution looking for a problem. Here are concrete applications we are deploying with clients:

1. 24/7 Customer Service That Does Not Sleep

The problem: Australian businesses lose customers to competitors who respond faster. After-hours calls go to voicemail. Weekends and public holidays mean missed opportunities.

The solution: An AI voice agent answers every call instantly, 24/7/365. It handles common questions, takes messages, books appointments, and escalates complex issues to humans.

Real impact: A Melbourne trades business deployed Heya and saw their average response time drop from 4 hours to 4 seconds. After-hours enquiry conversion increased by 340%.

2. Appointment Booking and Calendar Management

The problem: Staff spend hours each week playing phone tag to book appointments. No-shows cost money. Double-bookings create chaos.

The solution: Voice AI checks real-time availability, books appointments, sends confirmations, and handles rescheduling; all through natural conversation.

Real impact: A Sydney dental practice automated 78% of their appointment booking calls, freeing their front desk staff to focus on in-person patient care.

3. Outbound Campaigns at Scale

The problem: Following up with leads, conducting surveys, or sending appointment reminders via phone calls does not scale. A human can make perhaps 50 calls per day.

The solution: Batch calling enables hundreds or thousands of simultaneous outbound calls with personalised, natural-sounding voice interactions.

Real impact: A property management company now contacts all tenants about maintenance schedules in a single morning rather than over two weeks.

4. First-Line Technical Support

The problem: IT help desks are overwhelmed with repetitive questions. “How do I reset my password?” takes the same amount of agent time as complex issues.

The solution: Voice AI handles tier-1 support, walks users through common procedures, and escalates only the genuinely complex issues.

Real impact: An Australian SaaS company reduced human support ticket volume by 62% while maintaining customer satisfaction scores.

5. Voice-Enabled Internal Assistants

The problem: Field workers need information but cannot easily type while on a job site.

The solution: Voice assistants that workers can call to check inventory, log time, or access procedures hands-free.

Real impact: A construction company deployed voice AI for site foremen, reducing administrative overhead by 8 hours per week per foreman.

How Agentive and Heya Work Together

Agentive is an AI consultancy. We build AI agents and AI staff for businesses that want to operate faster, leaner, and smarter.

Heya is a specialised AI voice agent platform built for Australian businesses.

Together, we offer a complete solution:

Agentive provides:

- Strategic AI consulting to identify where voice AI delivers the highest ROI
- Integration with your existing business systems (CRM, booking systems, knowledge bases)
- Custom AI agent development for complex multi-channel workflows
- Ongoing optimisation and support from our Sydney-based team

Heya provides:

- Production-ready Australian voice agent platform
- Natural Australian voice with local understanding
- Enterprise-grade reliability and compliance
- Rapid deployment (often within days)

This partnership means you get the best of both worlds: strategic expertise from Agentive and specialised voice AI capability from Heya.

MyAgentive: Adding Voice to Your Personal AI Agent

For individual professionals and power users, MyAgentive (our open-source personal AI agent) already integrates with voice technologies including ElevenLabs and Deepgram.

Current capabilities: - Make phone calls with natural AI-generated voices via Twilio integration - Transcribe audio and video files with Deepgram - Generate voice content for podcasts, presentations, and more

Coming soon: - Full speech-to-speech conversation mode - Real-time voice interaction through MyAgentive's web and Telegram interfaces - Voice cloning for personalised assistant responses

MyAgentive is free, open-source, and runs on your own hardware. If you want to experiment with voice AI before committing to an enterprise deployment, MyAgentive is the perfect starting point.

Implementation Guide: Getting Started with Voice AI

If your organisation is considering voice AI, here is our recommended approach:

Step 1: Audit Your Voice Touchpoints (Week 1)

Map every place your business interacts with customers or staff via voice: - Inbound customer service calls - Appointment booking - Technical support - Sales enquiries - Internal communications

Identify which touchpoints have the highest volume, the most repetitive interactions, and the greatest impact when handled poorly.

Step 2: Choose Your First Use Case (Week 2)

Do not try to automate all voice interactions at once. Pick the single use case where: - Volume is high enough to justify automation - Interactions are relatively structured (booking, FAQs, simple enquiries) - Mistakes are recoverable (not high-stakes legal or medical advice) - Success is measurable (response time, resolution rate, customer satisfaction)

For most businesses, after-hours call handling or appointment booking is the ideal starting point.

Step 3: Define Success Metrics (Week 2)

Before deployment, establish how you will measure success: - Response time (seconds, not minutes) - Resolution rate (percentage of calls handled without human escalation) - Customer satisfaction (post-call surveys) - Cost per interaction - Staff time freed up

Step 4: Deploy and Iterate (Weeks 3-4)

Start with a pilot deployment. Monitor closely. Voice AI improves with feedback, so plan for an initial tuning period.

Step 5: Expand Strategically (Month 2+)

Once your first use case is working well, consider expanding to additional touchpoints. The integration work done for your first deployment makes subsequent deployments faster and cheaper.

Common Questions About Voice AI

Is voice AI ready for production use? Yes. OpenAI's Realtime API is out of beta. ElevenLabs has deployed Conversational AI 2.0 at enterprise scale. Heya is handling thousands of calls for Australian businesses daily. This is not experimental technology.

Will customers know they are talking to AI? They will likely guess, but the experience is dramatically better than traditional IVR systems. The key is setting appropriate expectations and ensuring easy escalation to humans when needed.

What about accents and dialects? Modern speech-to-speech systems handle accents well. Heya specifically trains on Australian speech patterns. For multilingual needs, platforms like Fish Speech and ElevenLabs support dozens of languages.

Is my data secure? Enterprise platforms like ElevenLabs offer SOC 2, HIPAA, and GDPR compliance. Data is encrypted in transit and at rest. Always verify compliance requirements with your chosen provider.

How much does it cost? Costs vary by platform and usage. OpenAI's gpt-realtime pricing is $32 per million audio input tokens and $64 per million output tokens. For comparison, that is roughly $0.03-0.06 per minute of conversation. Heya offers subscription pricing designed for small business budgets.

Can voice AI integrate with my existing systems? Yes. Modern platforms offer APIs and integrations with CRMs, booking systems, knowledge bases, and more. This is where Agentive adds value: connecting voice AI to your specific business systems.

The Future: What Is Coming Next

Speech-to-speech AI is advancing rapidly. Here is what we expect over the next 12-18 months:

Even lower latency: Sub-100ms response times will become standard, making AI conversations indistinguishable from human timing.

Improved emotion understanding: Models will better detect frustration, urgency, or confusion and adjust their responses accordingly.

Multilingual real-time translation: Speak in English, have your AI respond in Mandarin, Japanese, or any other language; in real time.

On-device processing: Models like NeuTTS Air are already running on phones. Expect more edge deployment options that work without internet connectivity.

Deeper system integration: Voice AI will move beyond conversation to taking action; booking, purchasing, updating records, and more; all through natural voice commands.

Key Takeaways

1. **Speech-to-speech AI is a fundamental shift, not an incremental improvement.** Single-model architectures that process audio directly deliver dramatically better latency, naturalness, and understanding than traditional pipelines.

2. **The technology is production-ready.** OpenAI, ElevenLabs, Heya, and others are serving real businesses today at scale.

3. **Australian businesses have local options.** Heya offers Australian voice, Australian understanding, and Australian support; no need to rely solely on US platforms.

4. **Start with one high-impact use case.** After-hours call handling or appointment booking are ideal starting points for most businesses.

5. **Integration matters.** The value of voice AI multiplies when it connects to your CRM, booking system, and knowledge base. This is where Agentive helps.

6. **MyAgentive users can experiment today.** Voice capabilities including phone calls, transcription, and voice generation are already available in our open-source personal AI agent.

Ready to Transform Your Customer Interactions with Voice AI?

Voice AI is no longer a future possibility. It is a present reality that leading Australian businesses are deploying now.

Whether you are a small business owner tired of missing after-hours calls, or an enterprise looking to transform your customer service at scale, Agentive and Heya can help.

Book a discovery call to explore how speech-to-speech AI can work for your organisation.

Book a Demo with Agentive: https://calendar.app.google/6Ue7azf2btWDFvks9

Agentive is an AI consultancy based in Sydney, Australia. We partner with Heya (heya.au), Australia's smartest AI voice agent platform, to help organisations deploy voice AI that actually works. Learn more at agentive.au.

Sources

- OpenAI: Introducing gpt-realtime and Realtime API updates (https://openai.com/index/introducing-gpt-realtime/)
- OpenAI: Voice agents documentation (https://platform.openai.com/docs/guides/voice-agents)
- ElevenLabs: Conversational AI 2.0 (https://elevenlabs.io/blog/conversational-ai-2-0)
- ElevenLabs: Voice agents and Conversational AI trends (https://elevenlabs.io/blog/voice-agents-and-conversational-ai-new-developer-trends-2025)
- Heya: Australia's Smartest AI Voice Agent (https://heya.au/)
- BentoML: Best Open-Source Text-to-Speech Models (https://www.bentoml.com/blog/exploring-the-world-of-open-source-text-to-speech-models)

- Northflank: Best Open-Source Speech-to-Text Models (https://northflank.com/blog/best-open-source-speech-to-text-stt-model-in-2026-benchmarks)