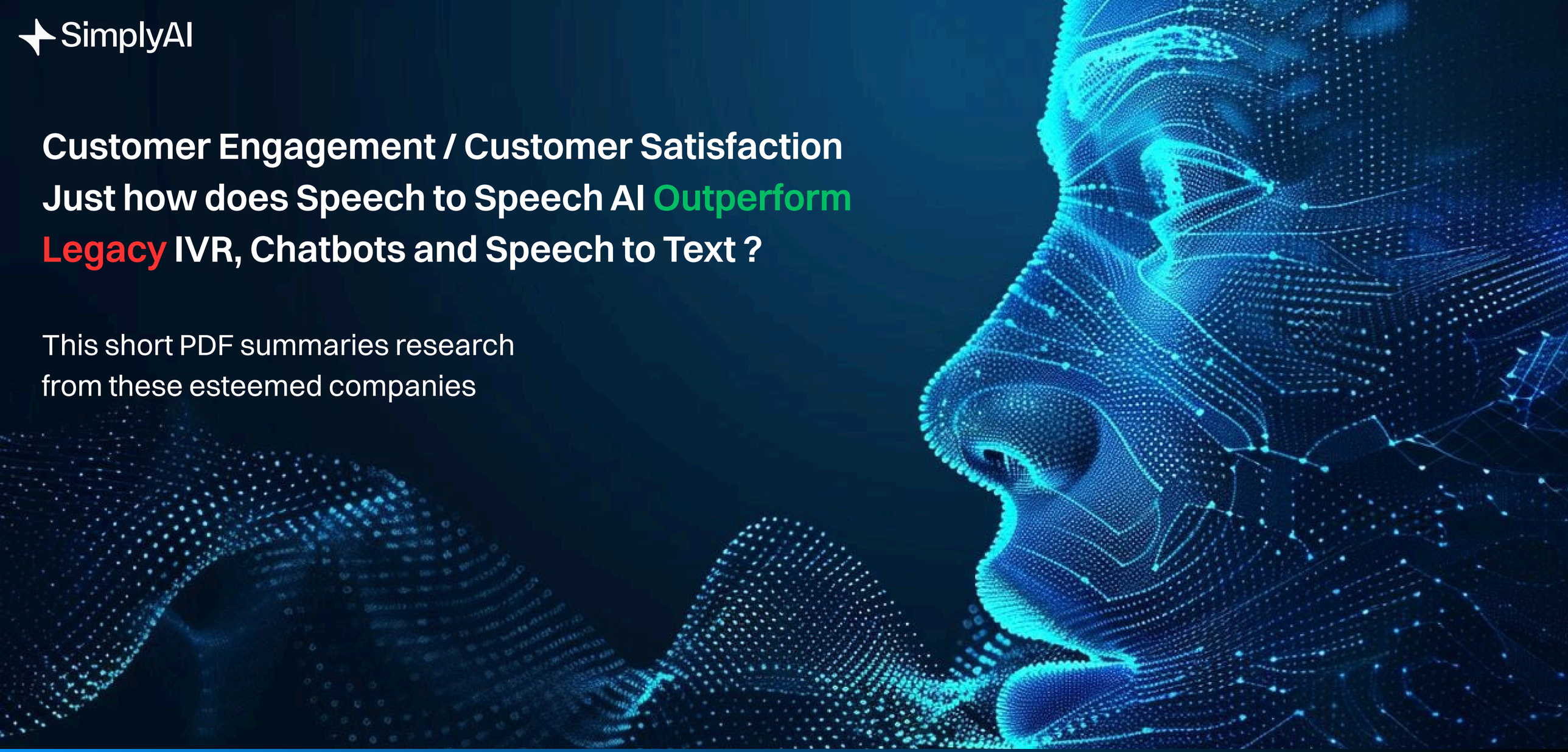
SimplyAI

Customer Engagement / Customer Satisfaction
Just how does Speech to Speech AI Outperform Legacy IVR, Chatbots and Speech to Text ?

This short PDF summaries research
from these esteemed companies



Gartner
FORRESTER

Why Speech-to-Speech AI Wins

Metric	IVR	Chatbot	Speech to Text	Speech to Speech AI
Containment Rate	40-50%	50-60%	✗	85%
Cust SAT	60-70%	65-75%	70%	80-90%
First Call Resolution (FCR)	70%	60%	✗	85-90%
Emotional Awareness	✗	✗	✗	✓
Multilingual Support	⚠ Manual	⚠ Manual	✗	✓

Suggested Messaging by Objection

Current Status / Mindset	Are you aware?
“Our IVR is already working fine”	Most IVR’s only contain 40% of calls. Speech to Speech AI Agents boosts this to 80%+ while improving customer satisfaction.
“This sounds expensive”	Start small - We offer pilot programs that show and ROI in weeks. See our ROI Reports of 10-40X returns v previous technologies
“We use chatbots, why add voice?”	Voice is still 60-70% of inbound support. This upgrades the core channel by +30%. Why have the cost of 30% of calls you do not need? Why not deliver a better, faster service?
“Won’t this replace agents?”	No - It frees them to handle higher-value conversations. Could more time to handle higher value calls improve KPI’s?
“Is this just speech-to-text?”	No - It’s full conversation: Voice in → Natural voice out. Real-time! No fewer than 10 technological advancements have delivered true speech to speech 6X faster to respond than speech to text.

Why Speech to Speech AI is the true IVR upgrade

Metric	Traditional IVR	Speech to Speech AI
Resolution Quality	Basic, Mechanical	Intelligent, Personalised
CX Impact	Flat, Menu Fatigue	Human-like Emotionally Aware
Efficiency Gains	Moderate	High ROI across all KPI’s
Future-Readiness	Legacy Tech	Conversational AI backbone

Bottom line: Speech to Speech AI turns your IVR from a static gatekeeper into a dynamic, human-like voice agent that listens, understands, and solves problems - in real time, at scale.

Benchmark Stats from Voice AI Departments

Metric	Traditional IVR	Speech to Speech AI	Source
Call Containment	40-45%	Up 70-85%	Deloitte, Genesys, NICE
Cust SAT Increase	-	+20-30 pts	Forrester CX Index, IBM Voice AI reports
Av Handling Time Reduction	-	Down 25% avg	CallMiner, Verint
Live Agent Deflection	-	Up to 60%	Intelligent, Personalised
IVR Abandonment	-20%	Down to 10-12%	Cognigy, NICE
Multilingual Support Impact	Limited	Engagement +35%	LivePerson, Unbabel

Speech to Speech AI V Chatbots

📊 Metric	🤖 Chatbots	🗣️ Speech-to-Speech AI	📌 Source
Call/Query Containment Rate	40–60%	✅ 70–85%	Deloitte, McKinsey, Cognigy
Customer Satisfaction (CSAT)	65–75%	✅ 80–90%	Forrester, Gartner CX Survey
Net Promoter Score (NPS)	Neutral to slightly positive	✅ +20 to +40 pts increase	NICE CXone, Genesys
Engagement Rate	Moderate (~40–50%)	✅ High (↑ 60–75%)	CallMiner, LivePerson
First Contact Resolution (FCR)	60–70%	✅ 85–90%	IBM Voice AI, Verint
Average Handle Time (AHT)	Reduced (vs human)	✅ 15–30% lower than chatbots	Contact Babel, Twilio
Abandonment Rate (self-service)	~20%	✅ ↓ to 10–12%	Genesys, NICE
Multilingual Support	Manual configs	✅ Built-in speech translation	Simply AI, Deepgram
Emotional Personalization	❌ None	✅ Responds empathetically	Cognigy, Kore.ai

Speech-to-Speech AI vs. Chatbots

Benefits & KPI Performance Comparison

Capability	Traditional Chatbots	Speech to Speech AI
Input Mode	Text only (Typed via web/app)	Spoken input (Natural voice)
Output Format	Text Response	Natural voice response (emotional, synthetic, multilingual)
Real-Time interction	No typing delays, scripted	Instant, conversational, dynamic
Experience Type	Menu-driven, FAQ-Style	Human-like, free-flowing dialogue
Use Case Fit	Web Chate, Self-service	Voice support, live calls, IVR, kiosks, retail
Emotional Detection	None	Detects tone, sentiment, urgency
Accessibility	No, requires typing, literacy	Hands-free, inclusive, voice-first

Performance Comparison on Engagement & Resolution

📈 KPI	Speech to Speech AI	Speech to Text	Source
CSAT (Customer Satisfaction)	✅ 80-90%	⚠ 60-70%	Forrester, NICE CX Index
Containment Rate	✅ Up to 85%	⚠ ~30-40%	Cognigly, IBM, Deloitte
First Contact Resolution (FCR)	✅ 85-90%	⚠ ~60-70%	Contact Babel, Verint
NPS (Net Promoter Score)	✅ +20 to +40 pts lift	⚠ Neutral impact	Genesys, Gartner
Average Handle Time (AHT)	✅ ↓ 20-30%	⚠ No direct reduction	Twilio, LivePerson
Abandonment Rate (IVR/Voice)	✅ ↓ to 10-12%	⚠ 20-25%	McKinsey, NICE CXone
Agent Workload Reduction	✅ High	⚠ Low	Cognigly, NICE

Differences between Speech to Speech v Speech to Text AI on Engagement & Resolution

🔍 Category	🗣️ Speech-to-Speech (S2S) AI	📝 Speech-to-Text (S2T) AI
Core Function	Converts voice to voice in real time using AI synthesis	Converts voice to written text
Input/Output Mode	Spoken → Spoken	Spoken → Text
Real-Time Engagement	✅ Yes – ideal for live conversations	⚠️ Often delayed – used post-call or asynchronously
User Experience	Conversational, fluid, natural	Functional, requires manual follow-up
Emotional Awareness	✅ Detects tone, emotion, and urgency	❌ Lacks tone/emotion detection
Multilingual Capability	✅ Real-time translation between spoken languages	⚠️ Requires add-on tools for translation
Use Case Fit	Voicebots, IVRs, live support, kiosks	Transcripts, notes, compliance, analytics
Agent Deflection	✅ High – automates live call flows	❌ Low – often used to assist agents, not replace them
First Contact Resolution	✅ High – handles queries live	⚠️ Relies on follow-up
Integration Complexity	Moderate (voice synthesis, latency sensitive)	Low (easy API plug-in)
Brand Experience	✅ Human-like voice reflects brand tone	❌ No brand personality