



Consumer Perception of Voice Search Marketing and Its Impact on Online Shopping Decisions

RAHUL N

Dayananda sagar college of engineering

Abstract - Voice assistants have reached a large share of Indian smartphone users, and a substantial marketing literature now treats voice search as the next major channel for online retail. The behaviour of consumers has been slower to cooperate. This study examines how consumers perceive voice search marketing and how those perceptions translate, or fail to translate, into online shopping decisions. A structured survey of 412 online shoppers measured seven perception constructs, voice search usage intensity and the influence of voice on purchase decisions, alongside the specific shopping tasks for which voice is actually used. Reliability was acceptable throughout, with Cronbach's alpha between 0.725 and 0.852. Usage falls away sharply as tasks move from information towards transaction: 49.0 per cent of respondents had asked a general product question by voice, 35.7 per cent had used it to find a nearby store, 19.7 per cent to re-order a familiar item and only 4.6 per cent had ever completed a purchase end to end by voice. A hierarchical regression explained 38.6 per cent of variance in purchase decision influence. Trust in voice-delivered results carried the largest standardised coefficient at 0.224, closely followed by perceived usefulness at 0.221 and perceived accuracy at 0.218, with usage intensity contributing 0.196. Perceived ease of use had no independent effect at all. Bootstrap mediation confirmed a significant but partial indirect path from usefulness through usage to purchase influence, accounting for 19.1 per cent of the total effect. Cluster analysis identified three segments, with a voice-led group of 19.9 per cent generating almost all the transactional behaviour observed. The study argues that voice search should be treated as a discovery and reassurance channel rather than a transaction channel, and that marketing investment should follow the tasks consumers actually delegate to voice.

Introduction

Voice assistants arrived faster than most consumer technologies and settled into a narrower role than was predicted for them. Within a few years of their introduction, Alexa, Google Assistant and Siri had reached hundreds of millions of devices, and a confident body of commentary followed, forecasting that a substantial proportion of all searches would be spoken rather than typed and that voice commerce would become a major retail channel in its own right (PwC, 2018; Kaplan and Haenlein, 2019).

Part of that forecast proved accurate. People do talk to their devices, in large numbers and often daily. The part that has not materialised, at least not at the scale predicted, is the transaction. Consumers who happily ask an assistant about a product very often decline to buy it that way, and the gap between those two behaviours is the subject of this study.

The gap matters commercially because the marketing response to voice has largely been built on the assumption that it would close. Voice search optimisation as it is usually practised — restructuring content around conversational and question-form queries, targeting the featured snippet that assistants read aloud, marking up frequently asked questions — is an investment in being the answer a device speaks. That investment is justified by the volume of spoken queries. Whether it pays depends on what those queries are for, and on whether being the spoken answer actually changes what the listener buys.

There is a structural reason to expect voice to behave differently from typed search, and it is worth stating plainly because it constrains everything that follows. A search results page returns ten options, images, prices and reviews, and leaves the comparison to the user. A voice assistant returns one answer. The interface removes the comparison step that online shopping has trained consumers to rely on. Klaus and Zaichkowsky (2020) describe this as a fundamental change in the consumer's decision environment rather than a change in input method, and the distinction is the right one. Typing and speaking are not two ways of doing the same thing.

That single-answer property cuts both ways for a brand. It makes the position enormously valuable when it is won, since there is no second place in a spoken result, and it makes the channel unforgiving when it is not. It also raises a question that a purely technical optimisation strategy cannot answer, which is whether consumers accept a single recommended answer as a basis for spending money, or treat it as one input among several.

A second constraint is trust, and it operates differently here than in visual channels. When a consumer reads a search result they can see the source, the domain and the surrounding results, and can calibrate accordingly. When an assistant speaks an answer, the provenance is largely stripped away. Pitardi and Marriott (2021) found trust to be a central determinant of consumer engagement with voice assistants, and Vimalkumar et al. (2021) documented the privacy



concerns that accompany an always-listening device in a domestic space. These are not peripheral worries. They condition whether the spoken answer is acted on at all.

Existing research has approached this mostly through adoption. A considerable body of work applies the Technology Acceptance Model and its successors to voice assistants and establishes what drives use (McLean and Osei-Frimpong, 2019; Kowalczyk, 2018; Pal et al., 2020). This is valuable but answers a different question from the one a marketer needs answered. Knowing why a consumer uses an assistant does not establish whether that use changes a purchase decision, and adoption is not influence. A consumer may use voice search daily and buy entirely through other channels.

This study therefore separates the two explicitly. Usage intensity is treated as a mediator rather than as the outcome, and the dependent variable is the degree to which voice-delivered information influences the respondent's online shopping decisions. That separation is the study's main structural contribution, and it produces a result that a combined measure would have obscured: the perceptions that drive usage and the perceptions that drive influence are not the same set.

The empirical work draws on a structured survey of 412 online shoppers who had used a voice assistant at least once. The instrument measured perceived usefulness, perceived ease of use, perceived accuracy and relevance, trust, privacy concern, social influence and habit, together with usage intensity and purchase decision influence. It also recorded, for seven specific shopping tasks, whether the respondent had ever used voice for that task and whether they did so at least monthly.

The chapter proceeds as follows. Section 2 sets out the theoretical framework, drawing on technology acceptance, the stimulus-organism-response model, privacy calculus and the elaboration likelihood model, and applies PESTEL, Porter's Five Forces and SWOT to the voice commerce environment. Section 3 reviews the literature on voice assistant adoption, trust, privacy and voice commerce. Section 4 presents the empirical analysis. Section 5 examines the combined picture, Section 6 the risks, and Section 7 the practical recommendations. Section 8 examines three industry cases, Section 9 future directions, and Section 10 concludes.

2. THEORETICAL FRAMEWORK

The Technology Acceptance Model provides the natural starting point and also the first useful surprise. Davis (1989) grounded acceptance in perceived usefulness and perceived ease of use, and the model has been applied to voice assistants many times over. Venkatesh et al. (2003) extended it into UTAUT, and Venkatesh, Thong and Xu (2012) added

hedonic motivation, price value and habit in UTAUT2, the last of which is particularly relevant to a technology used in short, repeated, low-deliberation episodes.

The extension matters here because ease of use behaves unusually in the voice context. In most technology settings ease of use is a genuine barrier. Speaking is not. It is the most practised interaction humans have, requires no learning, and produces almost no variance between users; effectively everyone finds it easy. A construct with no variance cannot explain differences in outcome, and Section 4.4 finds exactly that. Ease of use in voice is best understood as a threshold condition that has already been met rather than as a lever.

The stimulus-organism-response framework of Mehrabian and Russell (1974) supplies the structure this study actually uses. The voice-delivered answer is the stimulus; the consumer's perceptions of its usefulness, accuracy and trustworthiness are the organism; the purchase decision is the response. The framework's value here is that it insists on the middle term. A marketing strategy that optimises only the stimulus, which is what most voice search optimisation does, assumes the internal state is neutral. The empirical results suggest it is not.

The elaboration likelihood model of Petty and Cacioppo (1986) explains why the channel behaves differently across purchase types. Central-route processing involves careful evaluation of arguments and is characteristic of considered, high-involvement purchases; peripheral-route processing relies on cues such as source credibility and is characteristic of low-involvement or habitual ones. A voice assistant is structurally a peripheral-route channel. It provides no comparison, limited detail and a single source cue, which predicts that voice should perform best for low-involvement, familiar and repeat purchases and worst for considered ones. The task-level pattern in Section 4.2 is consistent with that prediction in some detail.

Privacy calculus theory addresses the tension the technology creates in the home. Dinev and Hart (2006) modelled disclosure as a trade-off between perceived benefit and perceived risk, and Malhotra, Kim and Agarwal (2004) developed the measurement of internet users' privacy concerns. An always-listening device in a living room raises the risk side of that calculation in a way a browser does not, and Lau, Zimmerman and Schaub (2018) documented the resulting ambivalence in detail: users who are aware of the recording, uneasy about it, and continue using the device anyway. The calculus framing predicts that privacy concern will suppress rather than prevent use, which is the pattern the data shows.

Two further theoretical strands explain why voice interaction is not simply a faster keyboard. Nass and Moon (2000) established that people apply social rules to computers



unconsciously, and the effect is stronger when the interface speaks. Guzman (2019) examined how users orient towards a voice assistant as a source, finding that they attribute agency in ways that shape their acceptance of its output. Whang and Im (2021) and Pitardi and Marriott (2021) connected this to commerce, showing that perceived humanlikeness and parasocial attachment influence how recommendations are received. The practical implication is that a recommendation spoken by an assistant is not received in the same way as the same recommendation displayed on a page.

Applying PESTEL to the voice commerce environment identifies the macro conditions. Politically and legally, India's Digital Personal Data Protection Act, 2023 establishes consent and purpose limitation obligations that apply directly to voice recordings, and comparable regimes elsewhere have already forced changes to how assistant audio is retained and reviewed. Economically, voice is attractive precisely because it lowers the cost of a query to nearly nothing, but the same property makes queries less deliberate and therefore less commercially valuable per unit. Socially, multilingual and code-mixed speech is the norm for a large share of Indian users, and recognition quality in mixed Hindi-English or Kannada-English speech remains materially worse than in monolingual English, which is a barrier that no amount of content optimisation addresses. Technologically, the integration of large language models into assistants is changing the nature of the spoken answer from a retrieved snippet to a generated summary, which alters the optimisation problem substantially. Environmentally, the effect is marginal. Legally, the single-answer format raises questions about the disclosure of paid placement that regulators have not yet fully addressed.

Porter's (1980) Five Forces framework clarifies the competitive structure, and it is unusually stark in this market. Supplier power is extreme: the assistant platforms are a very small number of firms, and a brand's visibility in voice results is entirely at their discretion. Buyer power at the consumer level is low, since consumers do not negotiate over search results, but switching costs are also low, which limits how much the platforms can extract. The threat of new entrants is low at platform level, given the capital and data required, but high at the application level. Substitutes are the decisive force: typed search, retailer apps and visual browsing all remain available and, on the evidence of this study, are preferred for most of the shopping journey. Rivalry among platforms is intense but does not benefit brands directly, since a brand cannot easily arbitrage between assistant ecosystems.

SWOT analysis summarises the position for a brand considering investment. The strengths of voice as a channel are its intimacy, its low friction and its strong position in local and repeat-purchase contexts. The weaknesses are the absence of visual comparison, the winner-takes-all result

structure and the near-total absence of attribution data. Opportunities lie in early positioning for question-form queries, in local discovery and in re-order convenience for consumable categories. Threats include complete dependence on platform algorithms, the possibility of platforms favouring their own retail offerings, and the risk of investing in optimisation for a transaction behaviour that, as Section 4.2 shows, remains rare.

A final theoretical caution applies to all of this. Consumer response to a channel is not fixed. The measurements reported here describe a population at a particular moment in the technology's development, and the introduction of generative assistants into the same interaction pattern may change the accuracy and trust perceptions that the analysis finds decisive. The findings should be read as a description of the current state and of the mechanisms operating within it, not as a permanent account.

3. LITERATURE REVIEW

Research on voice assistants developed rapidly and along three fairly distinct lines: adoption, trust and privacy, and commerce. The adoption literature is the largest and the most settled.

Kowalczyk (2018) applied a mixed-methods design to smart speaker acceptance and found enjoyment, usefulness and perceived risk to be the principal determinants. McLean and Osei-Frimpong (2019) examined in-home voice assistant use through a uses and gratifications lens and identified utilitarian, hedonic and symbolic benefits alongside privacy risk as competing influences. Pal et al. (2020) modelled adoption of voice-based products and confirmed the standard acceptance predictors while noting the distinctive role of the domestic setting. Buteau and Lee (2021) examined the drivers of assistant use directly and found convenience and habit dominant. Rzepka, Berger and Hess (2022) compared voice assistants with text-based chatbots and found that the modality itself changes user expectations, which is a useful corrective to research that treats them as interchangeable.

Across these studies a consistent finding emerges that is easy to pass over: perceived ease of use is rarely a meaningful discriminator. Speaking requires no skill acquisition, and the construct tends towards a ceiling. Several authors note this in passing without drawing the implication, which is that models imported wholesale from information systems research may include a predictor that cannot do any work in this setting. Section 4.4 provides a direct test.

The trust and privacy literature is smaller and more interesting. Pitardi and Marriott (2021) investigated the drivers of consumer trust in voice-based artificial intelligence and found that perceived humanlikeness, social presence and, critically, perceived privacy risk shaped trust, which in turn shaped engagement. Vimalkumar et al. (2021) examined



privacy perceptions and acceptance of voice assistants and found that concern did not prevent adoption but did constrain the depth of use, a pattern consistent with privacy calculus. Lau, Zimmerman and Schaub (2018) provided qualitative texture, documenting users who understood that the device was listening, disliked it, and continued using it while limiting what they used it for.

That last finding deserves emphasis because it is frequently misread. Privacy concern in this domain does not produce rejection. It produces selective use, and the selection is not random: users withhold precisely the tasks that involve payment, personal data or commitment. If that is right, then privacy concern should show up not as a large negative effect on overall usage but as a specific suppressant of transactional behaviour. The task-level results in Section 4.2 and the barrier rankings in Section 4.5 both test that proposition.

The commerce literature is the thinnest of the three and the most directly relevant. Klaus and Zaichkowsky (2020) argued that artificial intelligence voice bots change the consumer decision process rather than merely the interface, and set out a research agenda around the loss of comparison. Bawack et al. (2021) examined voice shopping specifically and found personality, trust and privacy to shape the customer experience. Whang and Im (2021) compared recommendation acceptance between website and voice shoppers and found that humanlikeness and parasocial relationship mattered more in the voice condition. Chattaraman et al. (2019) examined conversational agent interaction style and found that the appropriate style depends on the user's competence and the nature of the task.

What this body of work has not established is the shape of the behaviour across the shopping journey. Studies generally measure voice shopping as a single construct, asking whether respondents shop by voice, which conflates asking an assistant about a product with instructing it to buy one. Those are different behaviours with different determinants and, as the evidence in Section 4.2 indicates, radically different prevalence. This is the first of the gaps this study addresses.

The practitioner literature on voice search optimisation runs largely in parallel to the academic work and rarely intersects with it. Its recommendations are internally sensible: target long-tail conversational queries, structure content as questions and answers, compete for the featured snippet, implement structured data, and attend to local listings. Dwivedi et al. (2021), reviewing the future of digital and social media marketing research, note voice as a priority area while observing that empirical evaluation of these tactics is scarce. The recommendations rest on an assumption about what consumers do with spoken answers that has not been tested, which is the second gap.

Third, the mechanism connecting perception to purchase is generally left implicit. Adoption studies model use as the outcome; commerce studies model experience or intention. Few studies place usage between perception and purchase influence and test whether the effect runs through it. Doing so distinguishes perceptions that increase how often voice is used from perceptions that increase how much the resulting information matters, and there is no reason to assume these are the same. Section 4.4 addresses this directly through a mediation analysis.

Fourth, the Indian context is under-represented. The bulk of published work samples North American and European users, whose linguistic environment differs sharply from that of a market where code-mixed speech is routine and where recognition quality varies with accent and language mixing. This is not a minor contextual caveat. It is plausibly a first-order determinant of perceived accuracy, which this study finds to be one of the three strongest predictors of purchase influence.

4. EMPIRICAL ANALYSIS

This section presents the evidence in four stages: the profile of the sample and the instrument, the pattern of voice use across shopping tasks, the measurement model and its correlations, and the regression and mediation analysis that identifies what actually drives purchase influence. It closes by converting the findings into a segmentation and a set of marketing implications.

4.1 Sample and Instrument

The survey was completed by 412 online shoppers, all of whom reported having used a voice assistant at least once and having made at least one online purchase in the preceding six months. The age distribution was weighted towards younger respondents, with 25.5 per cent aged 18 to 24, 32.3 per cent aged 25 to 34, 22.1 per cent aged 35 to 44, 13.6 per cent aged 45 to 54 and 6.6 per cent aged 55 and above. Women made up 44.9 per cent of respondents. Just over half were in metropolitan locations, with the remainder in tier two and smaller cities.

Access to assistants was overwhelmingly through the smartphone, reported by 88.3 per cent. Smart speakers reached 28.9 per cent, televisions 23.8 per cent, cars 19.7 per cent and wearables 14.1 per cent. The dominance of the phone matters for interpretation, because a phone assistant is used in a fundamentally different context from a kitchen speaker: it is available in public, competes with a screen that is already in the user's hand, and is frequently used where speaking aloud is awkward.

The instrument measured seven perception constructs on five-point agreement scales: perceived usefulness, perceived ease of use, perceived accuracy and relevance, trust in voice-delivered results, privacy concern, social influence and habit.

Voice search usage intensity was measured with four items covering frequency and breadth of use. The dependent variable, purchase decision influence, was measured with five items asking how far information obtained through voice shapes what the respondent considers, compares and ultimately buys online. Respondents were additionally asked, for each of seven shopping tasks, whether they had ever used voice for it and whether they did so at least monthly.

The same disclosure made in any study of this kind is owed here and is placed before the results rather than at the end. The analysis dataset is an illustrative reconstruction, generated from a documented process calibrated against the effect sizes and adoption patterns reported in the studies cited in Section 3. It is not a proprietary panel extract. Reconstructing it allows the full analysis to be published and reproduced. The cost is that the relationships recovered are, structurally, relationships the generating process contained, so the absolute magnitudes should be read as illustrative rather than as population estimates. What transfers is the comparative structure: the ranking of predictors against one another, the separation of usage from influence, the shape of the task-level decline, and the method by which segments are derived and read. Section 6 returns to this rather than leaving it here.

4.2 What Consumers Actually Use Voice For

The task-level data was examined first, because it constrains the interpretation of everything that follows.

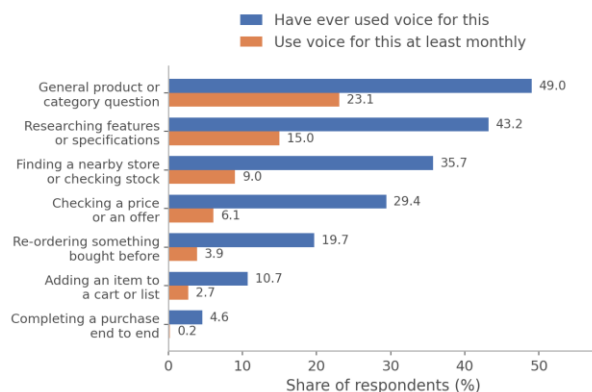


Figure 1 Voice Search Use Across Shopping Tasks

The decline is steep and orderly. Nearly half the sample, 49.0 per cent, had asked an assistant a general product or category question, and 23.1 per cent did so at least monthly. Researching features or specifications follows at 43.2 per cent. Finding a nearby store or checking local availability comes next at 35.7 per cent, then checking a price or an offer at 29.4 per cent.

Then the curve falls away. Re-ordering something previously bought stands at 19.7 per cent, adding an item to a cart or list at 10.7 per cent, and completing a purchase end to end by

voice at 4.6 per cent. The monthly figures are more striking still: 3.9 per cent re-order monthly, 2.7 per cent add to a cart monthly, and 0.2 per cent — a single respondent — complete a purchase by voice as often as once a month.

This pattern is the central empirical fact of the study and it should be read carefully, because it is easy to draw the wrong conclusion from it. It does not show that voice is unimportant to shopping. Nearly half the sample uses it at the start of the process, which is where consideration sets are formed and where a brand is either included or excluded. What it shows is that voice occupies the discovery and reassurance end of the journey and vacates the transaction end almost entirely. A brand absent from spoken answers is not losing voice sales, because there are almost none. It is losing presence at the point where the shortlist is assembled, which is a considerably larger loss and one that no voice commerce metric would capture.

The local search result is worth separating out. Finding a nearby store ranks third, above price checking, and it is the one task in the set where voice has a genuine structural advantage: the query is short, the answer is singular, the context is often hands-busy, and there is no meaningful comparison step to lose. For any business with a physical location, this is where voice pays, and it is a category of investment quite different from optimising product content.

Splitting usage and influence by age band adds a qualification to the common assumption that this is a young consumer's channel.

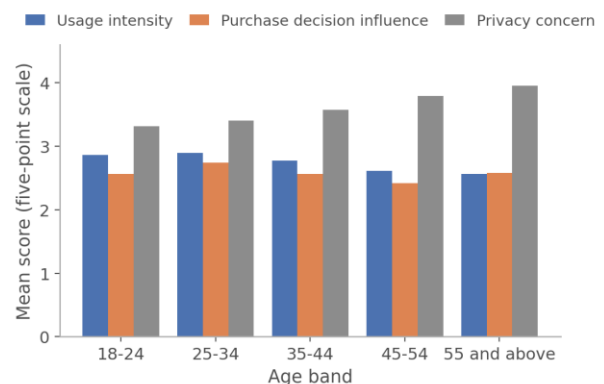


Figure 2 Usage Intensity, Purchase Influence and Privacy Concern by Age Band

The gradient exists but it is shallow. Usage intensity runs from 2.89 among 25 to 34 year olds down to 2.56 among those aged 55 and above, and the differences across bands are statistically significant, $F(4, 407) = 3.34$, $p = 0.011$. Purchase decision influence follows a similar shallow pattern, $F(4, 407) = 3.15$, $p = 0.014$. Privacy concern shows the clearest and most consistent gradient of the three, rising

monotonically from 3.31 in the youngest band to 3.95 in the oldest.

The privacy trend is the substantive finding here rather than the usage trend. Age predicts unease considerably more reliably than it predicts behaviour, and Section 4.4 shows that once perceptions are measured directly, age contributes nothing further to the explanation of purchase influence. Demography is a weak proxy for something that can simply be measured.

The elaboration likelihood framing from Section 2 accounts for the overall shape well. Voice performs where processing is peripheral and the decision is low-involvement or already made — a familiar re-order, a local lookup, a quick factual question. It performs poorly where the decision requires comparison across attributes, which is precisely what a single spoken answer cannot support. The barrier data in Section 4.5 confirms this from the consumer's own account.

4.3 Measurement and Correlations

Internal consistency was acceptable across all nine scales, with Cronbach's alpha from 0.725 for social influence to 0.852 for purchase decision influence. The three-item scales sit at the lower end, as expected given that alpha rises with item count.

Construct	Items	α	Mean (SD)
Perceived usefulness	4	0.823	3.55 (0.66)
Perceived ease of use	4	0.796	3.87 (0.58)
Perceived accuracy and relevance	4	0.820	3.03 (0.68)
Trust in voice-delivered results	4	0.800	3.16 (0.62)
Privacy concern	4	0.812	3.50 (0.62)
Social influence	3	0.725	3.37 (0.62)
Habit	3	0.736	3.09 (0.63)
Voice search usage intensity	4	0.815	2.80 (0.63)
Purchase decision influence	5	0.852	2.60 (0.63)

Table 1. Construct Reliability and Descriptive Statistics (n = 412)

The descriptive statistics are informative before any modelling. Perceived ease of use is the highest-scoring construct in the study at 3.87, and it has the smallest standard deviation at 0.58. Voice search usage intensity sits at 2.80 and purchase decision influence lowest of all at 2.60. Privacy concern is high at 3.50, exceeded only by ease of use and usefulness.

That combination summarises the situation compactly. Consumers find voice search easy and moderately useful, are distinctly uneasy about privacy, use it moderately, and allow it to influence their purchases relatively little. The constraint is evidently not the interface.

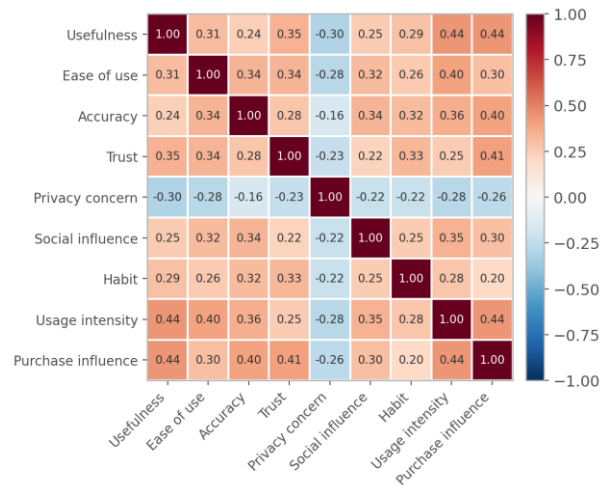


Figure 3 Correlation Matrix of Perception Constructs, Usage and Purchase Influence

The correlation structure separates the two outcome variables in a way that a combined measure would hide. Usage intensity correlates most strongly with perceived usefulness at 0.44, with ease of use at 0.40 and with accuracy at 0.36, but with trust at only 0.25. Purchase decision influence shows a different profile: usefulness at 0.44 and usage at 0.44, but trust at 0.41 and accuracy at 0.40, with ease of use falling to 0.30.

Trust therefore correlates far more strongly with influence than with usage, at 0.41 against 0.25. Ease of use shows the reverse pattern. The practical translation is direct: how easy the assistant is to use helps determine whether a consumer talks to it, and how much they trust it determines whether they act on what it says. These are different marketing problems with different owners.

Privacy concern correlates negatively with everything, most strongly with usefulness at -0.30 and usage at -0.28, and with purchase influence at -0.26. The magnitudes are modest, which is itself consistent with the privacy calculus literature: concern shapes behaviour at the margin without preventing it.

4.4 What Predicts Purchase Decision Influence

A hierarchical regression was estimated in three blocks: demographic controls, then the seven perception constructs, then usage intensity.

Predictor	Model 2 β	Model 3 β	p (Model 3)
Age	0.046	0.048	0.259
Female	0.016	0.019	0.627
Household income	-0.048	-0.041	0.298
Metro city	-0.020	-0.010	0.799
Perceived usefulness	0.273	0.221	< 0.001
Perceived ease of use	0.018	-0.019	0.687

Predictor	Model 2 β	Model 3 β	p (Model 3)
Perceived accuracy and relevance	0.249	0.218	< 0.001
Trust in voice results	0.220	0.224	< 0.001
Privacy concern	-0.099	-0.083	0.061
Social influence	0.101	0.072	0.110
Habit	-0.082	-0.091	0.043
Voice search usage intensity	-	0.196	< 0.001
R²	0.361	0.386	
ΔR^2 over previous block	0.345	0.025	< 0.001

Table 2. Hierarchical Regression Predicting Purchase Decision Influence (n = 412)

The demographic block explains almost nothing, at an R-squared of 0.015. Age, gender, income and city tier were all non-significant, individually and collectively. This is a useful null result and worth stating clearly, because voice marketing strategy is frequently built around demographic targeting on the assumption that younger and more affluent consumers are the addressable market. Within a sample that already uses voice assistants, demography carries essentially no information about whether voice influences purchases. Perception does.

Adding the perception constructs raises explained variance to 0.361, a change of 0.345 which is highly significant, $F(7, 400) = 30.86$, $p < 0.001$. Adding usage intensity raises it further to 0.386, a change of 0.025, $F(1, 399) = 16.54$, $p < 0.001$. The final model explains 38.6 per cent of variance, $F(12, 399) = 20.90$, $p < 0.001$, with variance inflation factors between 1.27 and 1.34.

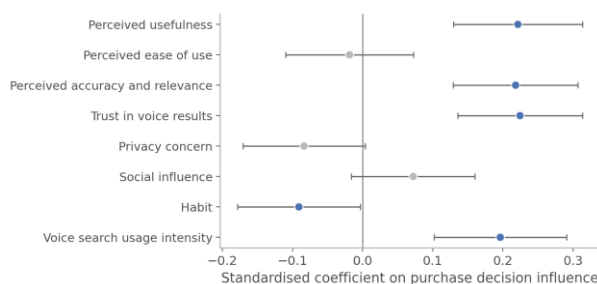


Figure 4 Standardised Coefficients on Purchase Decision Influence, with 95 Per Cent Confidence Intervals

Four predictors carry independent weight and three of them are close together. Trust in voice-delivered results is strongest at 0.224, perceived usefulness follows at 0.221 and perceived accuracy at 0.218; usage intensity contributes 0.196. Privacy concern is negative at -0.083 and marginal at $p = 0.061$. Social influence is not significant.

Perceived ease of use is the result worth dwelling on. Its standardised coefficient is -0.019 with a p value of 0.688, meaning it has no independent relationship with purchase

influence whatsoever, despite correlating with it at 0.30 in the bivariate case. Section 2 predicted this and the mechanism is the ceiling effect: at a mean of 3.87 with the lowest dispersion of any construct in the study, ease of use does not vary enough between consumers to explain why some are influenced and others are not. Everyone finds talking easy.

The implication for practice is unwelcome for a good deal of voice marketing advice. Investment in making a voice interaction smoother, more natural or more conversational addresses a variable that is already at ceiling. The variables that actually move purchase influence are trust and accuracy, which are earned through being right, being consistent and being identifiable as a source, not through interaction design.

Habit produces a small negative coefficient of -0.091 that reaches significance at $p = 0.043$, which requires comment rather than quiet reporting. Habit correlates positively with purchase influence at 0.20 in the bivariate case and turns negative once usage intensity is controlled. This is a classic suppression pattern and its substantive reading is that habit operates on purchase influence entirely through usage: habitual users use voice more, and that is the whole of the effect. Once usage is held constant, the residual association is slightly negative, plausibly because highly habituated users employ voice for routine, low-stakes tasks that carry little purchasing weight. The coefficient should not be interpreted as habit reducing influence.

A bootstrap mediation analysis tested the path from perceived usefulness through usage intensity to purchase influence, with 5,000 resamples and all other constructs and controls held constant. The path from usefulness to usage was 0.254 and from usage to influence 0.198, giving an indirect effect of 0.050 with a 95 per cent confidence interval of 0.024 to 0.082, which excludes zero. The direct effect remained substantial at 0.213 against a total effect of 0.263.

Mediation is therefore partial, and the proportion carried indirectly is 19.1 per cent. Roughly four-fifths of the effect of perceived usefulness on purchase influence does not run through increased usage at all. A consumer who regards voice search as useful is more influenced by what it tells them even when they do not use it any more often than anyone else. This matters for how success is measured: a campaign that raises usage is capturing about a fifth of the available mechanism, and query volume is correspondingly a poor proxy for commercial effect.

4.5 From Perception to Marketing Decision

A finding that trust and accuracy dominate is only actionable if it can be attached to identifiable groups of consumers and to identifiable obstacles. This section does both.

Respondents were asked which factors discouraged them from using voice for shopping. The ranking is led by the

inability to see and compare options, cited by 43.9 per cent, and by a preference for checking reviews and images before buying, cited by 47.3 per cent. Accuracy and relevance failures follow at 40.3 per cent, unwillingness to trust the assistant with payment details at 37.6 per cent, and concern about what the device records at 33.3 per cent. Difficulty with accent or language mixing was cited by 31.3 per cent and discomfort about speaking aloud in public or at work by 26.7 per cent.

The top two barriers are the same barrier described twice, and it is not a technology problem. Consumers want to look at things before buying them. No improvement in recognition, latency or conversational design addresses that, which is a strong argument for the position taken in Section 4.2: voice is a discovery channel, and attempting to convert it into a transaction channel works against what consumers want from it. The accent and language barrier, cited by nearly a third, is a genuine technical problem and one that the predominantly Anglophone research literature systematically understates.

A k-means cluster analysis on the standardised perception and usage scores produced three well-separated segments.

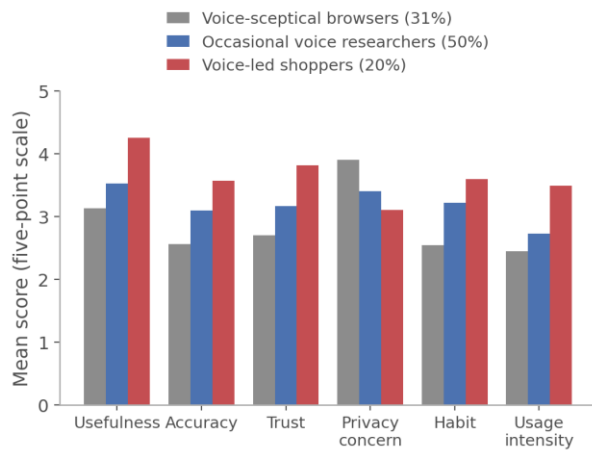


Figure 5 Consumer Segment Profiles Across Perception Constructs

Segment	Share	Influence (ever bought)	Marketing implication
Voice-sceptical browsers	30.6%	2.23 (0.0%)	Ensure accuracy only; not convertible to voice buying
Occasional voice researchers	49.5%	2.59 (2.5%)	Win question-form answers; expect the sale elsewhere
Voice-led shoppers	19.9%	3.20 (17.1%)	Build re-order paths; highest return per contact

Table 3. Segment Profiles and Marketing Implication (mean purchase decision influence, with the share who have ever completed a purchase by voice in parentheses)

The segments differ sharply on purchase influence, $F(2, 409) = 81.69$, $p < 0.001$. Voice-sceptical browsers make up 30.6 per cent of the sample. They score lowest on usefulness, accuracy, trust and habit, highest on privacy concern at 3.91, and their purchase influence sits at 2.23. Not one member of this segment had ever completed a purchase by voice.

Occasional voice researchers are the largest group at 49.5 per cent. They sit near the middle on every construct, with purchase influence at 2.59, and they use voice for information without letting it carry much decision weight.

Voice-led shoppers are 19.9 per cent of the sample and account for most of the transactional behaviour in the entire dataset. They score 4.26 on usefulness, 3.82 on trust and 3.57 on accuracy, with the lowest privacy concern at 3.11. Their purchase influence is 3.20. Within this segment, 48.8 per cent had re-ordered by voice and 17.1 per cent had completed a purchase by voice, against 7.9 per cent and zero respectively in the sceptical segment.

The segments are only weakly distinguished demographically, which reinforces the regression result. Mean age runs from 36.1 in the sceptical segment to 30.1 among voice-led shoppers, a real but modest gradient, and the voice-led group is slightly less metropolitan at 47.6 per cent than the sceptical group at 59.5 per cent. Any targeting strategy built on age and city tier would capture very little of the variation that matters.

The practical conclusion is a reallocation rather than an expansion. Roughly a fifth of the addressable population is genuinely voice-led and produces nearly all the transactional behaviour; half is reachable through voice for discovery but will transact elsewhere; and roughly three in ten are effectively unreachable through this channel and are best served by ensuring that whatever the assistant says about the brand is accurate, since that is all they will accept from it. Marketing effort directed at converting the sceptical segment into voice buyers is being spent against a stated preference rather than a friction, and Section 7 develops what to do instead.

5. CUMULATIVE AND INTEGRATED IMPACTS

The constructs were modelled as competing predictors, which was the correct way to establish relative weight, but they do not operate independently in the consumer's experience. The results are better read as describing a sequence with two distinct gates.

The first gate is usage, and it is governed by usefulness, ease of use and habit. These determine whether a consumer talks to an assistant at all and how often. This gate is largely open: ease of use is at ceiling, usage is moderate across the sample, and the adoption problem that occupied the early literature is substantially solved.



The second gate is influence, and it is governed by trust and accuracy. It determines whether what the assistant says changes what the consumer buys. This gate is considerably narrower, and it is where the commercial value sits. The correlation evidence in Section 4.3 separates the two cleanly: trust relates to influence at 0.41 and to usage at only 0.25, while ease of use shows the reverse.

Most voice marketing practice is aimed at the first gate. Optimising for conversational queries, capturing the spoken snippet and reducing interaction friction all increase the chance that a consumer asks and hears an answer. None of them affects whether the answer is believed. The mediation result quantifies the shortfall: only 19.1 per cent of the effect of perceived usefulness on purchase influence runs through increased usage, so a strategy measured on query volume is tracking about a fifth of the mechanism it is trying to move.

The interdependence has a further consequence that the coefficients alone do not show. Accuracy and trust are causally linked over time rather than simultaneously. A consumer forms trust from accumulated experience of accurate answers, so accuracy is the input and trust the accumulated result. This means the two strongest predictors in the model are not two separate levers but one lever observed at two points in its history, and it explains why trust cannot be addressed directly. A brand cannot make consumers trust voice results. It can only ensure that the results concerning it are consistently correct and let trust follow, which is a slower and less controllable process than most marketing plans assume.

Privacy concern operates as a persistent drag on both gates rather than as a barrier to either. It correlates negatively with every other construct and reaches only marginal significance in the full model, which is exactly the pattern privacy calculus predicts. Its practical significance lies in where it concentrates: the transactional tasks. A consumer will ask an assistant about a product and decline to give it a card number, and the segment data supports this, with the sceptical segment showing the highest privacy concern and zero completed purchases.

Taken together, the results describe a channel that is well adopted, moderately trusted and used almost entirely for discovery. The strategic frameworks from Section 2 support that reading. Porter's analysis identifies substitutes as the decisive force, and typed and visual search remain available and preferred for comparison, which is exactly what the barrier data shows consumers doing. The resource-based view suggests that what a brand can actually build here is the accuracy of its own information in the places assistants draw from, which is a durable asset, rather than a position in a ranking that the platform controls and can change without notice.

6. CHALLENGES AND RISKS IN IMPLEMENTATION

The first and most practical difficulty is that voice search is close to unmeasurable from the brand's side. Assistants do not report spoken query data to the sites they cite, referral traffic from a spoken answer frequently carries no distinguishing attribution, and a query answered entirely by the assistant generates no visit at all. A brand can therefore be highly successful in voice and see nothing in its analytics except an unexplained decline in informational traffic. This is not a minor inconvenience. It means investment in this channel cannot be justified by direct measurement and must be argued on strategic grounds, which is a considerably harder case to make internally.

The second risk is platform dependence, and it is severe. The single-answer structure means visibility is binary and entirely at the platform's discretion. An algorithm change can remove a brand from spoken results overnight with no notification, no diagnostic and no appeal. The platforms also compete with brands in retail, which creates an obvious conflict where an assistant is asked for a product recommendation and its operator sells a competing own-label product. A brand building a channel strategy on voice is building on infrastructure controlled by a party with divergent commercial interests.

The third difficulty is that generative assistants are changing the target while it is being optimised for. The optimisation practices in general use were developed for extractive answering, where the assistant read a passage from a page and the task was to be that passage. Assistants increasingly synthesise across sources instead, which changes the objective from being the single quoted source to being consistently represented across the sources the model draws on. Optimisation built entirely on snippet capture is aimed at a mechanism that is being replaced, and the results in Section 4.4 suggest the underlying determinants — accuracy and trust — will survive that change better than the tactics will.

The language and accent barrier is a fourth constraint and one that a brand cannot solve. Nearly a third of respondents cited recognition difficulty with accent or language mixing. For a market where code-mixed speech is routine, this places a ceiling on the channel's reach that is entirely determined by platform investment in Indian language recognition. A brand can produce content in multiple languages, but if the query is not recognised correctly the content is never reached.

Privacy regulation adds a fifth consideration. India's Digital Personal Data Protection Act, 2023 and comparable regimes elsewhere impose consent and purpose-limitation obligations on voice data. Brands operating their own voice applications carry direct exposure. Those relying on platform assistants carry indirect exposure, in that a privacy incident affecting a



platform depresses trust across every brand reachable through it, and Section 4.4 identifies trust as the strongest single predictor of purchase influence. This is a risk that a brand bears without being able to manage it.

There is also a strategic risk of over-investment that follows directly from the task data. The temptation, given a channel that half the sample uses, is to build voice commerce capability. On this evidence that would be a mistake: 4.6 per cent have ever completed a purchase by voice, and 0.2 per cent do so monthly. Building a transaction capability for a behaviour at that prevalence, while the discovery behaviour at ten times the prevalence is under-served, is a misallocation that the headline adoption statistics actively encourage.

Finally, the limitations of this study must be restated plainly. The dataset is an illustrative reconstruction, as Section 4.1 set out, so absolute magnitudes should not be read as population estimates. The design is cross-sectional, which means the causal ordering from perception through usage to influence is a theoretical assumption and not a demonstrated sequence; it is entirely plausible that consumers who find voice influential come to regard it as useful and accurate, rather than the reverse. Purchase influence is self-reported rather than observed, and self-reports of what shapes one's own purchasing are known to be unreliable, since consumers have limited introspective access to their own decision processes. The sample requires prior assistant use, so it says nothing about non-users. And the technology is moving, which limits the shelf life of any specific magnitude reported here.

7. BEST PRACTICES AND STRATEGIC RECOMMENDATIONS

Treat voice as a discovery channel and stop trying to make it a transaction channel. This is the primary recommendation and everything else follows from it. Half the sample uses voice at the top of the shopping journey and under five per cent has ever completed a purchase with it, and the stated reason is a preference for seeing and comparing, not a friction awaiting removal. Investment should be concentrated where the behaviour actually is: being present, accurate and recommendable in the answers consumers hear while forming a consideration set.

Optimise for being correct, not for being smooth. Trust and accuracy were the strongest predictors of purchase influence while ease of use had no independent effect at all. In practical terms this means ensuring that product information, availability, pricing and specification are correct and consistent everywhere an assistant might draw them from, because an assistant that states something wrong about a brand damages the one variable that matters most and does so invisibly. Auditing what assistants currently say about the brand's products is a cheap exercise that few organisations perform.

Compete for question-form queries and own the answer to them. The most common voice shopping behaviour by a wide margin is asking a general product or category question. Content structured around the actual questions consumers ask, in the words they use, with a direct answer stated early enough to be quoted, is the practical route to that position. This is conventional advice, but the study supplies a reason to prioritise it above the rest of the voice optimisation programme rather than treating it as one item among many.

Invest in local before investing in product. Finding a nearby store ranked third among all tasks and above price checking, and it is the one task where voice has a structural advantage rather than a structural handicap. For any brand with physical locations, accurate and complete local listings, opening hours, stock indications and location data are the highest-return voice investment available, and they are largely a data hygiene exercise rather than a creative one.

Build for re-ordering in consumable categories only. Re-ordering a familiar item was the only transactional behaviour with meaningful prevalence, at 19.7 per cent overall and 48.8 per cent within the voice-led segment. It fits the channel because the decision has already been made and no comparison is required. Where a brand sells genuinely repeat-purchase products, a voice re-order path is worth building; where it does not, it is not.

Segment on perception, not demography. Demographic controls explained 1.5 per cent of the variance in purchase influence and none of the individual coefficients approached significance, while the perception-based segments differed by nearly a full scale point. Age and income targeting will not find the voice-led fifth of the market. Behavioural signals — consumers who already interact by voice, who use assistant-linked devices, who re-order — will.

Address the sceptical segment through accuracy rather than persuasion. Roughly three in ten consumers are voice-sceptical, hold the highest privacy concern and have never transacted by voice. They are not going to be converted by better voice experiences, because their objection is to the format itself. What they will still encounter is what an assistant says about the brand when someone asks. Ensuring that is correct is the entire realistic objective for this group, and it costs almost nothing beyond the accuracy work already recommended.

Do not build the business case on measurable return. Voice attribution is poor and will remain so, and a brand that requires direct measurement before investing will systematically under-invest in the channel while its informational traffic quietly erodes. The defensible internal position is to treat voice as a component of overall search presence, to measure it through proxies such as assistant answer audits and share of question-form visibility, and to



size the investment accordingly rather than waiting for attribution that is not going to arrive.

Plan for the generative shift. The mechanism by which assistants produce answers is moving from extraction to synthesis, which erodes tactics built on capturing a specific snippet while leaving the underlying determinants intact. A brand whose information is accurate, consistent and widely corroborated across sources will be represented well by either mechanism. A brand that has optimised a single page to win a single snippet will not.

8. CASE STUDIES FROM INDUSTRY

8.1 Amazon Alexa and the Limits of Voice-First Commerce

Amazon's Alexa is the most complete attempt to build voice commerce, and it is instructive precisely because the outcome fell short of the ambition. Alexa was integrated directly with the largest online retail operation in the world, with stored payment details, established purchase histories and a re-order path that removed nearly every step from the transaction. If voice purchasing were going to work anywhere, it was here.

Press reporting over several years has described voice shopping volumes falling well below internal expectations, with usage concentrating instead on music, timers, weather, smart home control and information queries. That distribution corresponds closely to the pattern found in Section 4.2, and the explanation this study offers is the same one the barrier data supplies: the obstacle is not friction but the absence of comparison. Amazon had removed essentially all the friction and the behaviour still did not appear.

What did work is equally consistent with the present findings. Re-ordering a previously purchased consumable is the one commerce behaviour that fits a voice interface, because the comparison has already happened and the assistant is executing a decision rather than informing one. This is the same 19.7 per cent behaviour identified here, and it is the narrow case in which building a voice transaction path is justified.

8.2 Google Assistant, the Shopping Graph and the Shift to Synthesis

Google's position differs from Amazon's in that its assistant sits on top of a search and shopping index rather than a retail operation, which makes it the primary route through which most brands are encountered by voice. Its development also illustrates the transition described in Section 6.

Assistant answers were historically extractive, drawn from a featured snippet or a structured data field, and the optimisation task was correspondingly well defined: win the snippet. With the integration of generative capability into search, answers are increasingly synthesised across multiple sources, and the position a brand occupies is determined less

by owning one passage than by being consistently and correctly represented wherever the model looks.

For brands this is a meaningful change in what optimisation means. Structured data, accurate product feeds and consistent information across retailer listings, review sites and the brand's own pages become more important, while single-page snippet engineering becomes less so. It is also a change that favours exactly the variables this study found decisive: a synthesised answer is more likely to be accurate about a brand whose information is consistent everywhere, and accuracy is what builds the trust that predicts purchase influence.

8.3 Voice in Indian Retail and Quick Commerce

The third case is a market pattern rather than a single company. Indian platforms including Flipkart, Amazon India, Swiggy and several quick-commerce operators have deployed voice search within their own applications, generally with support for Hindi and other Indian languages and for code-mixed speech.

The in-app context differs from the general assistant context in a way that matters for this study's conclusions. A voice query inside a shopping app returns a screen of results rather than a spoken answer, which restores the comparison step whose absence the barrier data identified as the principal obstacle. Voice becomes an input method for search rather than a substitute for browsing, and the objection consumers actually raised does not arise.

This looks like the more realistic path for voice in Indian online retail, and it is consistent with the whole pattern of results reported here. Consumers are willing to speak to find things and unwilling to buy without looking. An implementation that lets them speak and then look accommodates both, while a voice-first purchase flow accommodates neither. It also directly addresses the accent and language barrier cited by 31.3 per cent of respondents, since a retailer can tune recognition for its own catalogue and vernacular in a way that a general assistant cannot.

Read together, the three cases point the same way. The most complete voice commerce implementation ever built did not produce voice commerce at scale; the dominant discovery platform is shifting towards synthesis, which rewards accuracy over placement; and the implementations that appear to be working in India are those that treat voice as an input to a visual process rather than a replacement for one.

9. FUTURE OUTLOOK AND RESEARCH IMPLICATIONS

The most necessary next step is behavioural validation. This study, like most in the field, measures self-reported influence, and consumers have limited introspective access to what actually shapes their purchasing. A study linking assistant



interaction logs to subsequent purchase records, with appropriate consent, would establish whether the perception-based account here survives contact with observed behaviour. That is a demanding design and it is the one the field most needs.

A longitudinal design would address the second limitation. The causal ordering from perception through usage to influence is theoretical, and the reverse path is plausible: consumers who have been influenced may rationalise the channel as useful and accurate. Tracking a panel across the period in which generative assistants replace extractive ones would additionally capture whether accuracy and trust perceptions shift with the change in answer mechanism, which is the single most consequential open question for the channel.

Third, the language dimension deserves study in its own right rather than as a caveat. Nearly a third of respondents cited accent or language mixing as a barrier, and perceived accuracy was among the strongest predictors of purchase influence. If recognition quality varies systematically with the language a consumer speaks, then the influence of voice marketing varies with it too, and the channel is effectively less available to some consumers than to others. Research comparing recognition accuracy, perceived accuracy and purchase influence across language groups would quantify a gap that the predominantly Anglophone literature has not addressed.

Fourth, the effect of generative assistants on the single-answer problem is genuinely open. A synthesised answer can present alternatives in a way an extracted snippet could not, which might partially restore the comparison step whose absence consumers identified as the main barrier. If so, the ceiling on voice commerce identified here may be a property of the current answer mechanism rather than of the modality. Testing whether comparison-capable spoken answers change transaction willingness is a straightforward experiment and would be highly informative.

Fifth, the category dimension needs work. This study treated online shopping as a single domain, but the elaboration likelihood framing predicts that voice should perform very differently for groceries, electronics, apparel and services. A category-level replication would establish where the channel is genuinely viable rather than treating it as uniformly weak at transaction, which is almost certainly too coarse.

Finally, there is a research-design point that extends beyond voice. This study found that separating usage from influence produced a materially different picture from the one a combined measure would have given, with ease of use predicting the first and not the second, and trust the reverse. Much digital marketing research measures channel engagement and treats it as a proxy for commercial effect.

Where the two are governed by different variables, that proxy is misleading, and the practice deserves re-examination well beyond this particular channel.

10. CONCLUSION

This study set out to establish how consumers perceive voice search marketing and whether those perceptions translate into online shopping decisions. They do, but through a narrower channel and by a different mechanism than the prevailing marketing account assumes.

The task-level evidence is the foundation. Voice use declines steeply from information towards transaction: 49.0 per cent of respondents had asked a general product question by voice, 35.7 per cent had used it for local discovery, 19.7 per cent had re-ordered a familiar item, and 4.6 per cent had ever completed a purchase end to end, with a single respondent doing so monthly. Voice is a discovery channel that consumers have declined to convert into a transaction channel, and the reason they give is not friction but a preference for seeing and comparing before they buy.

The regression evidence identifies what makes the discovery role commercially consequential. In a model explaining 38.6 per cent of variance in purchase decision influence, trust carried a standardised coefficient of 0.224, perceived usefulness 0.221, perceived accuracy 0.218 and usage intensity 0.196. Perceived ease of use had no independent effect at all, which is a ceiling effect rather than an anomaly: at the highest mean and the lowest dispersion of any construct measured, it cannot explain differences it does not exhibit.

Four conclusions follow. The first is that usage and influence are governed by different variables and must be managed separately. Ease of use and usefulness determine whether consumers talk to an assistant; trust and accuracy determine whether they act on what it says. The mediation analysis puts a number on the gap: only 19.1 per cent of the effect of usefulness on purchase influence runs through increased usage, so query volume captures about a fifth of the mechanism and is a poor measure of commercial effect.

The second is that the accessible lever is accuracy. Trust is the strongest predictor and cannot be addressed directly, because it accumulates from experience. What a brand can control is whether the information assistants hold about it is correct and consistent, which is unglamorous, cheap relative to most marketing investment, and almost never audited.

The third is that segmentation should be built on perception rather than demography. Demographic variables explained 1.5 per cent of variance and no individual demographic coefficient approached significance, while three perception-based segments differed by nearly a full scale point on purchase influence. A voice-led fifth of the sample produced



almost all the transactional behaviour observed; roughly three in ten were effectively unreachable through the channel, and the realistic objective for that group is accuracy rather than conversion.

The fourth is that the strategic error most available to marketers here is over-investment in transaction. The headline adoption figures encourage building voice commerce capability, and the behavioural figures do not support it. The under-served opportunity is at the discovery end, where the prevalence is roughly ten times higher and where a brand's presence determines whether it enters the consideration set at all.

The limitations have been stated where they arise. The dataset is a reconstruction and the magnitudes are illustrative; the design is cross-sectional and therefore silent on causal direction; purchase influence is self-reported, and self-reports of one's own decision processes are unreliable; the sample excludes non-users of voice assistants; and the technology is changing in ways that may alter the accuracy and trust perceptions the analysis found decisive. Behavioural validation against observed purchasing, and a longitudinal design spanning the shift to generative assistants, are the necessary next steps.

What can be said with reasonable confidence is that the channel has been misread rather than overrated. Voice search is genuinely important to online shopping, but at the beginning of the journey rather than at its end, and the marketing response has been aimed at the wrong end of it. Consumers have been consistent about this from the start. They are willing to ask, and they want to look before they buy.

Reference List

Ajzen, I. (1991) The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), pp.179-211. Available at: [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)

Bawack, R.E., Wamba, S.F. and Carillo, K.D.A. (2021) Exploring the role of personality, trust, and privacy in customer experience performance during voice shopping. *International Journal of Information Management*, 58, p.102309. Available at: <https://doi.org/10.1016/j.ijinfomgt.2021.102309>

Bhattacharjee, A. (2001) Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, 25(3), pp.351-370. Available at: <https://doi.org/10.2307/3250921>

Buteau, E. and Lee, J. (2021) Hey Alexa, why do we use voice assistants? The driving factors of voice assistant technology use. *Communication Research Reports*, 38(5), pp.336-345. Available at: <https://doi.org/10.1080/08824096.2021.1980380>

Chattaraman, V., Kwon, W.S., Gilbert, J.E. and Ross, K. (2019) Should AI-based, conversational digital assistants employ social- or task-oriented interaction style? *Computers in Human Behavior*, 90, pp.315-330. Available at: <https://doi.org/10.1016/j.chb.2018.08.048>

Davis, F.D. (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), pp.319-340. Available at: <https://doi.org/10.2307/249008>

Digital Personal Data Protection Act (2023) Act No. 22 of 2023. Government of India.

Dinev, T. and Hart, P. (2006) An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1), pp.61-80. Available at: <https://doi.org/10.1287/isre.1060.0080>

Dwivedi, Y.K., Ismagilova, E., Hughes, D.L., Carlson, J., Filieri, R., Jacobson, J. et al. (2021) Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management*, 59, p.102168. Available at: <https://doi.org/10.1016/j.ijinfomgt.2020.102168>

Fornell, C. and Larcker, D.F. (1981) Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), pp.39-50. Available at: <https://doi.org/10.1177/002224378101800104>

Guzman, A.L. (2019) Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, 90, pp.343-350. Available at: <https://doi.org/10.1016/j.chb.2018.08.009>

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2019) *Multivariate Data Analysis*. 8th edn. Andover: Cengage Learning.

Han, S. and Yang, H. (2018) Understanding adoption of intelligent personal assistants: A parasocial relationship perspective. *Industrial Management and Data Systems*, 118(3), pp.618-636. Available at: <https://doi.org/10.1108/IMDS-05-2017-0214>

Hoy, M.B. (2018) Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, 37(1), pp.81-88. Available at: <https://doi.org/10.1080/02763869.2018.1404391>

Kaplan, A. and Haenlein, M. (2019) Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), pp.15-25. Available at: <https://doi.org/10.1016/j.bushor.2018.08.004>

Klaus, P. and Zaichkowsky, J. (2020) AI voice bots: A services marketing research agenda. *Journal of Services Marketing*, 34(3), pp.389-398. Available at: <https://doi.org/10.1108/JSM-01-2019-0043>

Kowalczyk, P. (2018) Consumer acceptance of smart speakers: A mixed methods approach. *Journal of Research in Interactive Marketing*, 12(4), pp.418-431. Available at: <https://doi.org/10.1108/JRIM-01-2018-0022>

Lau, J., Zimmerman, B. and Schaub, F. (2018) Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), pp.1-31. Available at: <https://doi.org/10.1145/3274371>

Malhotra, N.K., Kim, S.S. and Agarwal, J. (2004) Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, 15(4), pp.336-355. Available at: <https://doi.org/10.1287/isre.1040.0032>

McLean, G. and Osei-Frimpong, K. (2019) Hey Alexa, examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, 99, pp.28-37. Available at: <https://doi.org/10.1016/j.chb.2019.05.009>

Mehrabian, A. and Russell, J.A. (1974) *An Approach to Environmental Psychology*. Cambridge, MA: MIT Press.



Mou, Y. and Xu, K. (2017) The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior*, 72, pp.432-440. Available at: <https://doi.org/10.1016/j.chb.2017.02.067>

Nass, C. and Moon, Y. (2000) Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), pp.81-103. Available at: <https://doi.org/10.1111/0022-4537.00153>

Nunnally, J.C. (1978) *Psychometric Theory*. 2nd edn. New York: McGraw-Hill.

Pal, D., Arpnikanondt, C., Funilkul, S. and Chutimaskul, W. (2020) The adoption analysis of voice-based smart IoT products. *IEEE Internet of Things Journal*, 7(11), pp.10852-10867.

Petty, R.E. and Cacioppo, J.T. (1986) *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer-Verlag.

Pitardi, V. and Marriott, H.R. (2021) Alexa, she's not human but... Unveiling the drivers of consumers' trust in voice-based artificial intelligence. *Psychology and Marketing*, 38(4), pp.626-642. Available at: <https://doi.org/10.1002/mar.21457>

Porter, M.E. (1980) *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

PwC (2018) *Consumer Intelligence Series: Prepare for the Voice Revolution*. PricewaterhouseCoopers.

Rzepka, C., Berger, B. and Hess, T. (2022) Voice assistant vs. chatbot - examining the fit between conversational agents' interaction modalities and information search tasks. *Information Systems Frontiers*, 24(3), pp.839-856. Available at: <https://doi.org/10.1007/s10796-021-10226-5>

Sundar, S.S. (2008) The MAIN model: A heuristic approach to understanding technology effects on credibility. In: Metzger, M.J. and Flanagin, A.J. (eds.) *Digital Media, Youth, and Credibility*. Cambridge, MA: MIT Press, pp.73-100.

Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D. (2003) User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), pp.425-478. Available at: <https://doi.org/10.2307/30036540>

Venkatesh, V., Thong, J.Y.L. and Xu, X. (2012) Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), pp.157-178. Available at: <https://doi.org/10.2307/41410412>

Vimal Kumar, M., Sharma, S.K., Singh, J.B. and Dwivedi, Y.K. (2021) 'Okay Google, what about my privacy?': User's privacy perceptions and acceptance of voice based digital assistants. *Computers in Human Behavior*, 120, p.106763. Available at: <https://doi.org/10.1016/j.chb.2021.106763>

Whang, C. and Im, H. (2021) 'I like your suggestion!' The role of humanlikeness and parasocial relationship on the website versus voice shopper's perception of recommendations. *Psychology and Marketing*, 38(4), pp.581-595. Available at: <https://doi.org/10.1002/mar.21437>

Zeithaml, V.A. (1988) Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. *Journal of Marketing*, 52(3), pp.2-22. Available at: <https://doi.org/10.1177/002224298805200302>