

Voice and Visual Search in the Consumer Buying Journey: An Empirical Study of Urban Shoppers in Hyderabad

Dr Mohd Moinuddin Mudassir,

Associate Professor, Shadan Institute of Computer Studies, Hyderabad

Abstract

The growth of interaction technologies powered by artificial intelligence (AI) has changed the way consumers discover, assess, and buy products. Voice search (which uses natural language processing) and visual search (which makes use of computer vision) brought mark a major departure from earlier text-based query systems. This study looks at the effects and level of adoption of voice and visual search technologies throughout the five stages of the consumer buying process: Need Recognition, Information Search, Evaluation of Alternatives, Purchase Decision, and Post-Purchase Behavior. A structured survey was carried out with a sample of 250 urban consumers in Hyderabad, Telangana, and the research combines the Unified Theory of Acceptance and Use of Technology (UTAUT) with the Consumer Decision-Making Model. Empirical analysis, carried out using multiple regression and paired-sample t-tests, shows that visual search has a greater influence at the Information Search and Alternative Evaluation stages, while voice search mainly affects Need Recognition and repeat Purchase Execution. The study also offers strategic recommendations for retail marketers regarding how to improve their multimodal search infrastructure in rapidly growing metropolitan areas.

Keywords: Voice Search, Visual Search, Consumer Buying Journey, Information Search, UTAUT, Urban Retailing, Hyderabad.

1. Introduction

Digital commerce is going through a shift in its underlying infrastructure from text-based input methods to multimodal, conversational, and image-based query systems. The combination of voice assistants (such as Google Assistant, Amazon Alexa, and Apple Siri) with visual recognition services (such as Google Lens, Pinterest Lens, and Instagram Visual Search) has changed the way consumers interact with e-commerce systems.

In places like Hyderabad in Telangana—where there is high ownership of smartphones, 5G network coverage is expanding and the population is digitally literate—multimodal search tools are now changing from being just innovative services to becoming essential intermediaries in the process of shopping. Earlier consumer decision-making models, for example the well-known Engel-Kollat-Blackwell (EKB) model, placed great emphasis on making active text queries and navigating directly to websites. Nowadays, however, voice and visual search decrease the mental effort required and reduce friction in transactions, thus altering consumer behaviour at every stage of the buying process.

Although voice search makes use of speed, the convenience of being able to operate it without having to use your hands, and conversational natural language processing (NLP), visual search employs image recognition in order to connect offline aesthetic inspiration with online product matching. It is important to understand the different effects of these two technologies at various stages of the consumer decision process if omnichannel marketing strategies, SEO optimization, and retail interface design are to be successfully implemented.

1.1 Research Objectives

1. To evaluate the adoption levels and user perceptions of voice and visual search technologies among urban retail consumers in Hyderabad.
2. To analyze the differential impact of voice search versus visual search across the five stages of the consumer buying journey.

3. To determine the key drivers (e.g., Performance Expectancy, Effort Expectancy, Perceived Risk) influencing the intention to use voice and visual search for commerce.
4. To provide strategic implications for retail marketers and e-commerce platforms operating in Indian metropolitan hubs.

2. Literature Review & Theoretical Framework

2.1 Technology Adoption and the Consumer Buying Journey

The theoretical anchoring of this study synthesizes Venkatesh et al.'s (2003) Unified Theory of Acceptance and Use of Technology (UTAUT) with the classic Consumer Decision-Making Process (Howard & Sheth, 1969; Blackwell, Miniard, & Engel, 2006).

Author(s) & Year	Key Focus	Primary Findings	Relevancy to Present Study
Kropf et al. (2021)	Voice Commerce Adoption	Reduces search friction; faces visual feedback barriers.	Theoretical basis for Voice in Purchase/Need Recognition.
Flavián et al. (2019)	Visual Technology & UX	Enhances spatial presence; superior for hedonic categories.	Theoretical basis for Visual in Alternative Evaluation.
Mari & Pitardi (2020)	Parasocial Voice AI Trust	Warmth builds trust, but accuracy drives conversion.	Informs Performance Expectancy construct in UTAUT.
Yang et al. (2022)	Visual Search vs. Text	Bypasses "vocabulary gap" in complex aesthetic queries.	Validates Visual search dominance in Information Search.
Kowalczyk (2018)	Resistance to Voice Tech	Privacy and security fears hinder transactional adoption.	Justifies inclusion of Perceived Risk (PR) construct.
Mishra & Singh (2021)	Indian Omnichannel Behavior	High mobile readiness; fragmented utility across tools.	Contextualizes urban shopping behavior in Hyderabad.
Guha et al. (2024)	Multimodal Integration	Synthesizes voice and vision for friction-free retail.	Justifies examining Voice and Visual search together.

3. Hypotheses Development

To empirically evaluate the directional impact and functional differences of voice and visual search across the purchase funnel, six directional **Alternative Hypotheses (H1)** were formulated:

- **H1:** Voice search has a significantly higher positive impact on the *Need Recognition* stage compared to visual search.
- **H2:** Visual search has a significantly higher positive impact on the *Information Search* stage compared to voice search.
- **H3:** Visual search has a significantly higher positive impact on the *Evaluation of Alternatives* stage compared to voice search.
- **H4:** Voice search has a significantly higher positive impact on executing the *Purchase Decision* for routine, low-involvement products compared to visual search.

- **H5:** Voice search has a significantly higher positive impact on *Post-Purchase Support* activities compared to visual search.
- **H6:** Performance Expectancy, Effort Expectancy, Social Influence, and Facilitating Conditions significantly increase the overall *Behavioral Intention to Adopt Multimodal Search* for e-commerce, while Perceived Risk significantly reduces it.

4. Research Methodology

4.1 Sampling and Data Collection

The study conducted a cross-sectional survey focusing on urban smartphone users in Hyderabad, Telangana. A structured questionnaire was used and circulated online using a convenience sampling method. A screening question ensured that respondents had actively used either voice search or visual search for product search or purchasing within the preceding 6 months. A total of 250 valid responses were considered for empirical analysis ($N = 250$).

4.2 Measurement Model & Scale Calibration

The instrument utilized a 5-point Likert Scale (1 = Strongly Disagree to 5 = Strongly Agree). The constructs measured included:

- **UTAUT Drivers:** Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), Facilitating Conditions (FC), and Perceived Risk (PR).
- **Funnel Stages (1–5):** Perceived Effectiveness of Voice vs. Visual Search across Need Recognition, Information Search, Evaluation, Purchase, and Post-Purchase.

The reliability of the constructs was verified using Cronbach's Alpha, where all subscales exceeded the 0.70 threshold (α Voice = 0.864, α Visual = 0.881, α Buying Journey = 0.835).

5. Empirical Data Analysis & Results

5.1 Sample Demographic Profile ($N = 250$)

Demographic Variable	Category	Frequency (N)	Percentage (%)
Gender	Male	134	53.6%
	Female	116	46.4%
Age Group	18-25 Years	88	35.2%
	26-35 Years	112	44.8%
	36-50 Years	38	15.2%
	Above 50 Years	12	4.8%
Occupation	IT/Corporate Professional	118	47.2%
	Business/Self-Employed	42	16.8%
	Student	64	25.6%
	Academician	26	10.4%
Primary Usage Preference	Voice Search Dominant	68	27.2%

Visual Search Dominant	94	37.6%
Dual Multimodal User	88	35.2%

5.2 Comparative Hypothesis Testing Across the Consumer Buying Journey

To test the alternative hypotheses H1 through H5, paired-sample t-tests were conducted comparing mean perceived effectiveness scores between Voice Search (μ_{Voice}) and Visual Search (μ_{Visual}) across the 250 urban respondents.

Table 1: Hypothesis Evaluation Matrix (Paired t-Test Results, N = 250)

Hypotheses	Voice Mean	Visual Mean	Mean Difference	SD	t-value	p-value (<0.05)	Decision
H1: Voice > Visual in Need Recognition	4.12	3.45	-0.67	0.82	12.92	0.000*	Accepted
H2: Visual > Voice in Information Search	3.30	4.38	+1.08	0.76	22.45	0.000*	Accepted
H3: Visual > Voice in Alternative Search	2.95	4.42	+1.47	0.81	28.67	0.000*	Accepted
H4: Voice > Visual in Purchase Decision	4.05	3.20	-0.85	0.88	15.26	0.000*	Accepted
H5: Voice > Visual in Post-Purchase Support	3.88	3.10	-0.78	0.85	14.50	0.000*	Accepted

*Statistically significant at the 5% level ($p < 0.05$).

5.3 Regression Analysis: Drivers of Multimodal Adoption (H6)

A multiple linear regression model was built to evaluate H6, measuring the directional impact of the UTAUT drivers and perceived risk on overall **Behavioral Intention (BI)** to adopt multimodal search tools:

$$BI = \beta_0 + \beta_1 (PE) + \beta_2 (EE) + \beta_3 (SI) + \beta_4 (FC) + \beta_5 (PR) + e$$

Table 2: Multiple Regression Model Summary (N = 250)

Independent Construct	Unstandardized β	Standard Error (SE)	Standardized β	t-value	p-value (<0.05)	Hypothesis effect
(constant)	0.625	0.185	-	3.378	0.001	-
PE	0.382	0.045	0.412	8.488	0.000*	Positive
EE	0.274	0.052	0.268	5.269	0.000*	Positive
SI	0.112	0.048	0.105	2.333	0.020*	Positive
FC	0.148	0.050	0.138	2.960	0.003*	Positive
PR	-0.195	0.042	-0.182	-4.642	0.000*	Negative

Model Statistics: $R = 0.782$, $R^2 = 0.612$, Adjusted $R^2 = 0.604$, $F(5, 244) = 77.08$, $p < 0.001$.

Decision: **H6 is fully Supported.**

5.4 Key Findings & Discussion

- **Confirmation of Functional Divergence (H1, H4, H5):**

H1, H4, and H5 confirm that voice search functions primarily as a task-oriented utility. It excels at fast trigger moments, repeat buying, and post-purchase inquiries like tracking delivery status.

- **Confirmation of Visual Superiority (H2, H3):**

H2 and H3 confirm that visual search dominates aesthetic and cognitive exploration phases. Online buyers in Hyderabad leverage image recognition to find visually similar items, compare offline pricing with online, and overcome text-based keyword limitations.

- **UTAUT Validation (H6):**

Performance Expectancy ($\beta = 0.412$) appeared as the primary positive predictor of behavioral adoption. This shows that consumers use search AI primarily because it delivers tangible accuracy and speed gains over traditional text queries. Perceived Risk ($\beta = -0.182$) appeared as a key barrier, driven by privacy concerns regarding inaccurate camera matches and ambient voice monitoring.

6. Strategic & Managerial Implications

Based on the validated hypotheses, Indian e-commerce and q-commerce companies should implement the following targeted initiatives:

- Digital interface must integrate native image-upload option into product search bars, and simultaneously supporting voice search option for quick search.
- E-commerce platforms having visual-heavy categories like apparel, interior, footwear, etc. should optimize visual metadata and high-resolution product catalogs to gain visibility in camera-based search engines.
- Grocery platforms should develop voice-activated "quick reorder" integrations to capture top-of-funnel shoppers.

7. Conclusion, Limitations & Future Scope

This study validates that voice and visual search are complementary tools occupying distinct niches across the consumer buying journey. Visual search drives exploration and alternative evaluation, while voice search accelerates initiation and purchase execution.

8. Limitations and Future Scope

- **Geographic Scope:** The sample is drawn from urban Hyderabad; rural and semi-urban behavior across Telangana may vary based on linguistic diversity and internet literacy.
- **Product Category Moderation:** Future studies should test whether product involvement (e.g., high-involvement smartphones vs. low-involvement groceries) moderates the strength of H1 through H5.

9. References

- **Blackwell, R. D., Miniard, P. W., & Engel, J. F. (2006).** *Consumer Behavior* (10th ed.). Mason, OH: Thomson Business and Economics.
- **Flavián, C., Ibáñez-Sánchez, S., & Orús, C. (2019).** The impact of augmented reality, virtual reality and 360-degree devices on the user experience. *Journal of Business Research*, 100, 547–560.

- **Guha, S., Chatterjee, A., & Banerjee, R. (2024).** Multimodal fluidity in e-commerce: The convergence of voice and visual search interfaces. *Journal of Retailing and Consumer Services*, 76, 103550.
- **Howard, J. A., & Sheth, J. N. (1969).** *The Theory of Buyer Behavior*. New York, NY: John Wiley & Sons.
- **Hoy, M. B. (2018).** Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, 37(1), 81–88.
- **Kowalczyk, P. (2018).** Consumer acceptance of smart speakers: A mixed methods approach to understanding privacy risks and voice commerce. *Journal of Research in Interactive Marketing*, 12(4), 418–437.
- **Kropf, M., Baumgarth, C., & Schmidt, S. (2021).** Voice commerce: A systematic literature review and research agenda. *International Journal of Consumer Studies*, 45(4), 785–802.
- **Kumar, A., & Sharma, P. (2023).** Mobile visual search in retail apps: Impact on user engagement and purchase intention among Gen Z. *International Journal of Information Management*, 70, 102631.
- **Mari, A., & Pitardi, V. (2020).** How smart voice assistants can shape consumer-brand relationships: The role of parasocial interaction and perceived warmth. *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS)*.
- **Mishra, R., & Singh, A. (2021).** Omnichannel retailing in emerging markets: A study of technology readiness in Indian metropolitan cities. *Journal of Business Research*, 132, 45–57.
- **Qin, Y., Li, X., & Wang, J. (2021).** Visual search and social commerce: How image recognition bridges inspiration and evaluation. *Electronic Commerce Research and Applications*, 49, 101088.
- **Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003).** User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.
- **Yang, Z., Sun, Y., & Zhang, Y. (2022).** Bridging the vocabulary gap: The effectiveness of visual search in e-commerce for hedonic products. *Decision Support Systems*, 155, 113725.