Perceived Documentation Status from Visual and Audio AI-Generated Cues: A Multimodal Conjoint Study

Sudhamshu Hosamane[1], Carolina Coimbra Vieira[2], Thomas Davidson[1], Kiran Garimella[1]
[1]*Rutgers University*, [2]*Max Planck Institute for Demographic Research, Germany*

Conjoint experiments have become the standard tool in political and social science for studying multidimensional preferences [3]. In the domain of attitudes towards migrants, the canonical study by Hainmueller and Hopkins [7] demonstrated that Americans hold a broad consensus regarding immigrant preferences, favoring educated, English-speaking applicants with employment plans. Since then, conjoint designs have been widely used to study immigration preferences across countries and policy contexts [1–3, 6, 16]. However, most existing immigration conjoint experiments rely on text-based attribute profiles, even though real-world attributions of citizenship and documentation status are formed rapidly from visual and auditory cues such as race, age, attire, accent, and language fluency. Recent work has shown that visual cues substantially shape political and social evaluations [8], yet, to the best of our knowledge, no previous study has jointly manipulated visual and audio information to investigate attitudes toward migrants.

In this work, we present the generated stimuli and their validation for a subsequent multimodal experiment designed to estimate how facial and vocal cues causally shape perceptions of documentation status. Individuals form rapid holistic impressions from embodied appearance and speech [5, 14], and text-based descriptions may underestimate the discriminatory effects of visible traits [8]. Recent advances in generative AI now enable scalable multimodal experimentation by allowing researchers to generate realistic, attribute-controlled images and synthetic voices while maintaining experimental control. Building on this methodological frontier [12, 13], we develop a conjoint design that combines AI-generated portraits with synthetic voice clips to estimate how appearance and speech jointly shape perceptions of citizenship, legality, and deportability.

The conjoint design jointly manipulates a series of visual and auditory characteristics. As shown in Table 1, the visual attributes include race/ethnicity, gender, age, religion, and attire, while the audio attributes vary in accent and English fluency.

To generate the visual stimuli, we collected base face images from four established human face databases: the Chicago Face Database, the Face Research Lab London Set, the Hawai'i Face Database, and the DeZIM Picture Database [4, 9, 10, 15]. The first three datasets provide standardized face stimuli for research use, with stated permissions or licenses allowing use of the images in research contexts, including altered/adapted stimuli where permitted by their terms. For the DeZIM dataset, we obtained access and permission directly from DeZIM. We used all images in accordance with the respective dataset terms and conditions. From each dataset, we selected two reference faces representing relatively low and high perceived attractiveness, a common feature across all datasets, to increase variation in baseline appearance while maintaining realism.

Using these reference images, we developed a generative AI pipeline that systematically modified specific facial attributes while preserving facial identity and overall image consistency. We implemented the editing pipeline with Qwen-Image-Edit-2511 [11], an open-source image-editing model, and ran inference on an NVIDIA H100 GPU through Google Colab. The pipeline generated variations in age, attire, and visible religious markers, enabling controlled manipulation of socially salient characteristics across profiles. Figure 1 shows examples of a base face and edited stimulus derived from it. Audio stimuli were generated separately using ElevenLabs text-to-speech models (v3), which allowed us to manipulate accent and English fluency while standardizing speech content, pacing, and recording quality. The full stimulus library is available at **https://go.rutgers.edu/stimuli**.

Before conducting the main conjoint experiment, we will conduct a large-scale validation study to assess the fidelity of the attributes and the perceived realism of the generated multimodal stimuli. The validation study will include all generated profiles (i.e., images + audio clips), which will be independently evaluated by annotators recruited through CloudResearch Connect. Each participant will evaluate approximately 15 randomly assigned profiles, and each profile will be rated by roughly 15 independent annotators to ensure reliable measurement of perceived characteristics. For each profile, annotators will be asked to identify the perceived demographic attributes listed in Table 1 as well as a series of social perception measures commonly used in social cognition research (e.g., how trustworthy, attractive, dominant, threatening, competent, and warm each profile appears).

The validation step serves two purposes. First, it allows us to evaluate whether the intended experimental manipulations are consistently recognized by respondents. Second, they provide continuous perceptual measures that capture latent social evaluations associated with different combinations of visual and auditory traits. Stimuli that fail to meet predefined recognition or realism thresholds will be discarded and regenerated, ensuring that the final conjoint experiment relies on stimuli that are both realistic and accurately perceived across conditions. The validation survey is available at **https://go.rutgers.edu/validation**.

The conjoint design includes both single-profile and forced-choice evaluations. For each profile, participants answer a series of questions measuring perceived documentation status and support for enforcement actions. Specifically, respondents are asked: (i) "How likely do you think this person is a U.S. citizen?", (ii) "How likely do you think this person is a documented immigrant?", and (iii) "How likely do you think this person is an undocumented immigrant?", all measured on continuous rating scales. Participants are also asked: (iv) "Do you think this person should be deported?" and (v) "Do you think the government will deport this person?" These two questions are designed to distinguish between respondents' normative preferences and their expectations about state behavior.

In the forced-choice component, participants compare two profiles simultaneously and answer comparative questions designed to capture relative judgments. These include: (i) "Which of these two individuals is more likely to be undocumented?", (ii) "Which of these two individuals should be deported?", and (iii) "Which of

Attribute	Levels
Race/Ethnicity (image)	White, Black, Hispanic/Latino, Middle Eastern, South Asian, East Asian, Pacific Islander
Gender (image)	Male, Female
Age (image)	Young Adult (18–30), Adult/Middle-aged (30–60), Older Adult (60+)
Religion (image)	Muslim (visible religious markers, e.g., hijab for women and kufi for men), Non-Muslim
Attire (image)	Casual (gray t-shirt, jeans), Formal (navy blazer over white dress shirt, gray trousers), Service (navy polo with embroidered burger-chain logo, work pants), Labor (dusty orange hi-visibility vest over a faded work shirt, khaki work pants).
Accent (audio)	Standard American English, Foreign-accented English
Fluency (audio)	Fluent English, Limited English proficiency

Table 1: Conjoint Attributes and Levels

Figure 1: Example stimuli: one base face image (left) and two edited variants (center, right).

these two individuals do you think the government would deport?" Together, these outcomes allow us to estimate how visual and auditory cues jointly shape perceptions of citizenship, legality, and deportation.

The contribution of this project is threefold. Methodologically, it develops and validates a scalable pipeline to generate photorealistic, attribute-controlled multimodal stimuli using generative models, combined with a rigorous human verification loop. To our knowledge, this is the first immigration conjoint experiment to employ AI-generated visual and audio stimuli. Substantively, the study extends Flores and Schachter [6]'s concept of "social illegality" beyond text and beyond static imagery by estimating how embodied appearance and voice jointly shape attributions of undocumented status and deportation deservingness. Practically, the project demonstrates a cost-effective framework for multimodal stimulus generation that can be adapted to broader research on discrimination, stereotyping, and social cognition.

At the time of writing, the project is in the stimulus validation stage, with full survey fielding planned for summer 2026. At the conference, we will present preliminary estimates from both forced-choice and single-profile outcomes, alongside validation evidence for the generated visual and audio stimuli. The resulting design provides causal evidence on how appearance and voice jointly shape Americans' perceptions of citizenship, illegality, and deportability, advancing immigration research beyond text-only conjoint paradigms and establishing a new methodological framework for multimodal social-computing experiments.

References

[1] Kirk Bansak, Jens Hainmueller, and Dominik Hangartner. 2016. How Economic, Humanitarian, and Religious Concerns Shape European Attitudes toward Asylum Seekers. *Science* 354, 6309 (2016), 217–222.

[2] Kirk Bansak, Jens Hainmueller, and Dominik Hangartner. 2023. Europeans' support for refugees of varying background is stable over time. *Nature* 620, 7975 (2023), 849–854.

[3] Kirk Bansak, Jens Hainmueller, Daniel J Hopkins, and Teppei Yamamoto. 2021. Conjoint Survey Experiments. In *Advances in Experimental Political Science*, James Druckman and Donald P Green (Eds.). Cambridge University Press.

[4] Lisa DeBruine and Benedict Jones. 2017. Face Research Lab London Set. doi:10.6084/m9.figshare.5047666.v5

[5] Susan T Fiske, Amy JC Cuddy, Peter Glick, and Jun Xu. 2002. A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition. *Journal of Personality and Social Psychology* 82, 6 (2002), 878–902.

[6] René D. Flores and Ariela Schachter. 2018. Who are the "Illegals"? The Social Construction of Illegality in the United States. *American Sociological Review* 83, 5 (Oct. 2018), 839–868. doi:10.1177/0003122418794635

[7] Jens Hainmueller and Daniel J Hopkins. 2015. The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants. *American Journal of Political Science* 59, 3 (2015), 529–548.

[8] Alberto López Ortega and Marco Radojevic. 2025. Visual Conjoint vs. Text Conjoint and the Differential Discriminatory Effect of (Visible) Social Categories. *Political Behavior* 47, 1 (March 2025), 335–353. doi:10.1007/s11109-024-09953-7

[9] Debbie S. Ma, Joshua Correll, and Bernd Wittenbrink. 2015. The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data. *Behavior Research Methods* 47, 4 (2015), 1122–1135. doi:10.3758/s13428-014-0532-5

[10] Chanel Meyers, Maria M. Garay, and Kristin Pauker. 2024. Hawai'i Face Database: A Racially and Ethnically Diverse Set of Facial Stimuli. *Self and Identity* (2024). doi:10.1080/15298868.2024.2417664

[11] Qwen Team. 2025. Qwen-Image-Edit-2511. https://huggingface.co/Qwen/Qwen-Image-Edit-2511. Hugging Face model card.

[12] Nicole Schwitter. 2025. Using Artificial Intelligence to Generate Visual Vignettes in Factorial Survey Experiments. *Social Science Computer Review* (2025), 08944393251392916.

[13] Ankit Sisodia, Alex Burnap, and Vineet Kumar. 2025. Generative Interpretable Visual Design: Using Disentanglement for Visual Conjoint Analysis. *Journal of Marketing Research* 62, 3 (June 2025), 405–428. doi:10.1177/00222437241276736

[14] Alexander Todorov, Anesu N Mandisodza, Amir Goren, and Crystal C Hall. 2005. Inferences of Competence from Faces Predict Election Outcomes. *Science* 308, 5728 (2005), 1623–1626.

[15] Susanne Veit and Iniobong Essien. 2022. The DeZIM Picture Database: Faces. doi:10.34882/dezim.dpd-fac.f.1.2.0 Image dataset.

[16] Kirill Zhirkov. 2025. Americans' Perceptions About Immigrants from Different World Regions: Evidence from a Multinomial Conjoint Experiment. *Journal of Experimental Political Science* (2025), 1–11.