

WHITEPAPER

AI VOICE: UNLOCKING MULTILINGUAL SYNTHETIC STORYTELLING

BY SIMON AMBROSE



LOCARIA

THE RISE OF SYNTHETIC STORYTELLING

The use of generative voice technology in marketing is set to transform aspects of the advertising industry by enabling brands to produce high-quality, human-like audio content at scale while significantly reducing the costs and time associated with traditional studio recording.

Marketers are leveraging these tools to extend their global reach, and scaling scripts and stories across languages without needing a vast network of voice actors. They are also using AI to add narrative voiceovers to content that traditionally relied on supers, while unlocking new capabilities such as avatar lip-syncing and personalisation capabilities for programmatic audio.

The surge in interest and adoption is accompanied by a need for a coordinated sonic strategy, a reworking and retooling of audio production processes, and the hiring or upskilling of audio engineers with prompting and post-editing competencies. As with all emerging generative technology, marketers need to understand when and where it can be utilised, while navigating company specific AI policies and regulatory challenges emerging from the complexities of voice ownership, deep faking and voice cloning.

CONTINUOUS TOOL TESTING

The developments in generative audio require a rigorous and ongoing tool testing strategy. The technology is progressive, with frequent model updates to improve accents, prosody, emotional ranges, characters, product features, and varying performance for different language pairs.

ElevenLabs is currently unrivalled for pure audio and creativity, offering hyper-realistic, emotive voice cloning that is perfect for narrative storytelling and high-quality campaign voice overs. HeyGen shines in video localisation projects, translating content with mostly accurate lip-syncing, and Synthesia is a powerhouse in avatar-led videos, perfect for instructional content and longer form YouTube explainers.

New models are now able to adjust the prosody (rhythm and stress) to match local cultural norms like making a voice sound more enthusiastic for a US audience or more reserved for a Japanese audience.

Experimenting with these specific strengths rather than relying on a single solution, brands can ensure they are deploying the right technology to maximise impact across different channels and need to keep up to date on new feature development in an increasingly competitive auditory marketplace.

PROMPT ENGINEERING & EVOLVING SKILL SETS

The skills and competencies required for AI voice production differ significantly from traditional casting. These span interpersonal abilities and emotional intelligence to identify unique vocal textures, negotiations with talent, and coaching actors in real time to elicit a nuanced performance, while AI is primarily focussed on technical manipulation via prompt syntaxes and post-editing engineering workflows.

The move from the director's chair to the dashboard requires the mastering of prompt engineering and phonetics to iteratively adjust pitch, speed, and intonation curve by curve. At Locaria, we look for AI engineers with a good knowledge and background of traditional voice production with the ability to create detailed syntax inputs and tweak product settings, to sculpt the audio to draw on styles and mimic the nuances of human idiosyncrasies.

OUR MULTILINGUAL AUDIO ENGINEERING APPROACH

- ① When working on global campaigns at Locaria, we start with the master campaign asset, typically a locked script and a high-fidelity reference recording, which serves as the emotion and blueprint for all future variations.
- ② The script will then under-go transcreation, a collaborative human based localisation approach to adapt words, idioms, cultural nuances, rhyme and rhythm, with relevant local, regional or global stakeholders signing off on the direction using backtranslations.
- ③ Together with the producer, a decision is made on the tooling and sonic approach; for example, whether to clone the source voice, whether to generate a new unique character voice, or to select from the array of preset voices within the software.
- ④ Although in its adoption infancy, some brands will look to apply AI lip-syncing, ensuring the words spoken in the local languages are matched with the visemes (mouth movements), creating a highly engaging local brand experience.
- ⑤ The prompting engineer will create a number of outputs, which are then reviewed by a native speaker, to review inflection quality, abnormalities and pronunciation (e.g. phonetic errors such as, I 'Live' at home vs they are playing 'Live'), which can be engineered or re-promoted out.
- ⑥ Our audio engineer will mix the AI-generated voice sample, adjusting EQ, compression, and reverb, to seamlessly blend the synthetic vocal with the video's original tone and flow, before final QC before final delivery to platforms.

THE ADVANTAGES OF AI

GLOBAL CONSISTENCY

AI voice can excel at global localisation, allowing a brand's sonic identity to be instantly translated into dozens of languages while retaining a singular tonal character (e.g. a founder, mascot or brand ambassador), ensuring a cohesive brand experience across international markets.

SPEED, SCALE, AND EFFICIENCY

Integrating AI voice into advertising allows for unprecedented ability to achieve scalability and cost-efficiency. It allows marketers to bypass the logistical bottlenecks of traditional recording such as studio booking, actor availability, and re-recordings, enabling the production of audio days rather than weeks, and the removal of usage fees can drive significant commercial savings.

HYPER-PERSONALISATION AT SCALE

AI opens a new world where brands can dynamically generate thousands of audio variations tailored to specific usernames, locations, or behaviours, creating dynamic audio advertising that has been tested to significantly boost engagement and recall.

THE DISADVANTAGES OF AI

NUANCES AND DELIVERY

Despite its efficiency, AI voice technology often struggles to replicate the emotional depth and subtle nuance of a professional human voice actor, potentially leading to a 'flat' delivery that fails to resonate in high-stakes, emotional storytelling. For high empathy advertising - think healthcare or banking - the lack of micro-imperfections (breaths, slight hesitations) that signal 'honesty' in human speech can be lost, creating a sense of unease or distrust.

HOMOGENISATION

If competing brands utilise the same popular stock AI voices, they lose their distinct sonic identity, blending into the background noise of the market. We see this on social media, where brands are using generic voices such as TikTok's Jessie, Matilda and Sarah or overused clones such as Morgan Freeman. New models allow for thousands of generative voices to be produced, but they lack the training data for longer form use-cases.

SUITABILITY AND EXPECTATIONS

For certain briefs, manipulating AI to match the original objective and to stay consistent across multiple campaigns, can be a highly challenging task. It may require interactive prompting and revisions, which can lead to the reallocation of savings on human talent to prompting, engineering and technology tokens.

ETHICAL AND LEGAL COMPLEXITIES

As with other emerging AI technologies, there are concerns regarding copyright ownership and the rights of voice actors and changing regulations across markets. We work with brands to understand their internal policies and platforms to monitor their stances as they develop within the industry.

LOCARIA LOCARIA
LOCARIA LOCARIA
LOCARIA LOCARIA
LOCARIA LOCARIA
LOCARIA LOCARIA
LOCARIA LOCARIA
LOCARIA LOCARIA



www.locaria.com