

Llais: Welsh Translation And Lip-Sync For Video Production

A Feasibility Report
By Aled Jones
Y Pod Cyf

media
cymru × pdr.

Introduction

Overview

The Llais project represents a pioneering initiative aimed at exploring the potential impacts and ethical implications of utilising artificial intelligence (AI) for lip-syncing in Welsh to English video content. This project seeks to assess how AI technologies can be fine-tuned and packaged for various markets, thereby enhancing the accessibility and distribution of Welsh-language media globally.

Summary

The Llais project is positioned to drive significant advancements in the Welsh-language media landscape by harnessing AI technology responsibly and ethically. By addressing industry concerns and incorporating stakeholder feedback, Llais aims to enhance the accessibility and quality of Welsh content while promoting inclusivity and representation on a global scale. A comprehensive report detailing testing outcomes and lessons learned will be provided at the end of the initiative.

Objectives and Scope

The project is focused on the following objectives:

- Investigating the effectiveness, ethics, and cost-efficiency of AI-driven lip-syncing in Welsh-language video production.
- Providing opportunities for translators and dubbing artists to create new services tailored for training, corporate video markets, and television production companies.
- Engaging with industry stakeholders to understand concerns and insights related to AI in the creative sector.

Project Team

Project Lead

Aled Jones

Aled Jones serves as the director of Y Pod Cyf, specialising in Welsh-language digital initiatives. Throughout his career, he has managed and developed innovative projects that have contributed significantly to shaping the Welsh digital landscape.

Aled sought to investigate the potential impacts and ethical considerations of using AI lip-syncing tools for producing Welsh-to-English video content, aiming to make Welsh-language content accessible to broader audiences. With over twenty years of experience in the Welsh media industry, Aled was able to leverage his extensive network to engage industry stakeholders and gain insights into AI's role in the creative sector.



Technical Partner:

Robin Moore, Director at Shwsh and Innovation Consultant with Media Cymru, serves as a technical partner in the Llais project, offering AI expertise.

Timeline

Key

- PDR sessions
- Media Cymru and consortium
- Alacrity sessions
- Research



July

11.07 - Media Cymru Seed Fund 2024 - Launch event
25.07 - PDR Session 1

August

Market exploration - Identifying target output and potential clients/end-users

September

Industry Interviews

03.09 - PDR - Y Pod - Seed Session 2a - Research
17.09 - Alacrity: Andrew Tuite
30.09 - Filming actors for creation of testing content

October

02.10 - PDR - Y Pod - Seed Session 2b - Presenting Insights

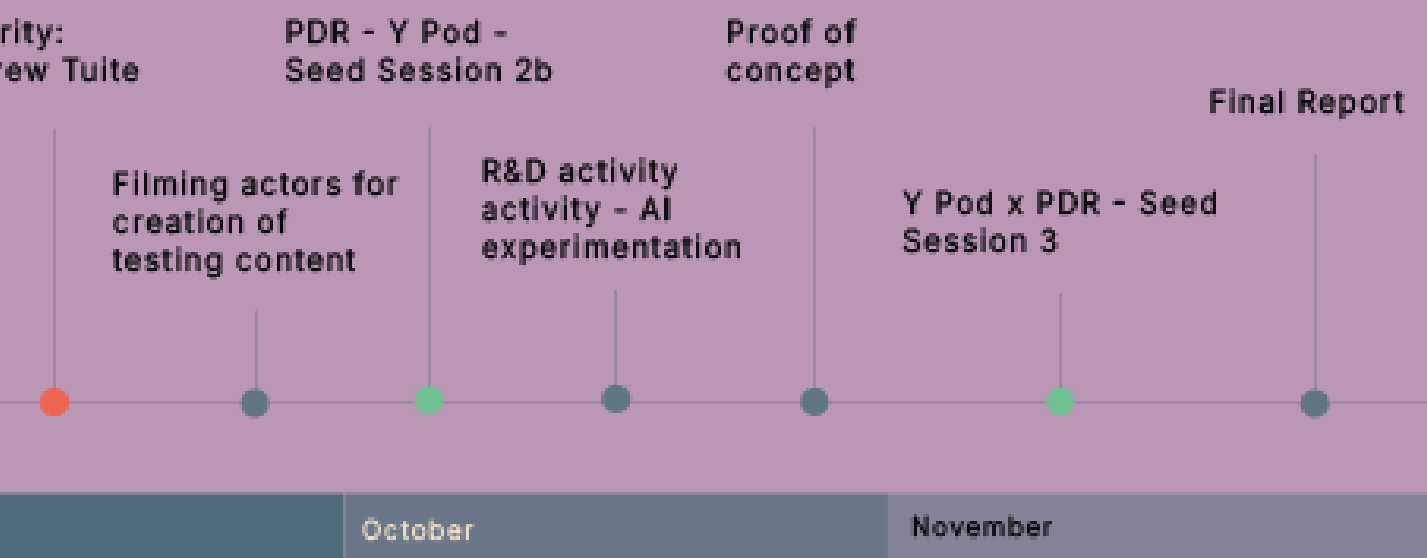
R&D activity activity - AI experimentation

Proof of concept

November

11.11 - Y Pod x PDR - Seed Session 3 (Final Session) - Research experience & report

27.11 - Final report: Comprehensive feasibility report compiling findings from all research activities, assessing the viability of the proposed solution, and providing recommendations for future development.





Project Definition

The objective of this project is to assess the potential impacts and ethical considerations of employing AI-based lip-syncing tools for the translation and production of Welsh-to-English video content. This investigation aims to understand how these tools could be optimised and packaged into a platform suitable for targeted markets, which will be identified as part of this research.



The Llais project seeks to examine the efficacy, ethical dimensions, and cost-effectiveness of using advanced AI lip-syncing technology in translating video content between Welsh and English. This project aims to create new service opportunities for translators and dubbing artists, particularly in the

corporate training and video production sectors, and to support television production companies in expanding their international distribution of Welsh-language content.

Llais is a project developed by Y Pod Cyf alongside technical AI partner Shwsh.

The Challenge

How might we explore the potential impacts and ethical implications of using AI lip syncing tools in the production of Welsh to English video content to scope how they might be fine-tuned and packaged to be offered as a platform to interested markets, that will be identified as part of this work.

Imagine if we could fine-tune these tools specifically for Welsh-to-English

translation and offer them as a platform for translation agencies or dubbing artists, enabling them to deliver enhanced services to their clients.

We explored both the potential market for this model and the quality of the outputs it could achieve.

Opportunities for Exploration



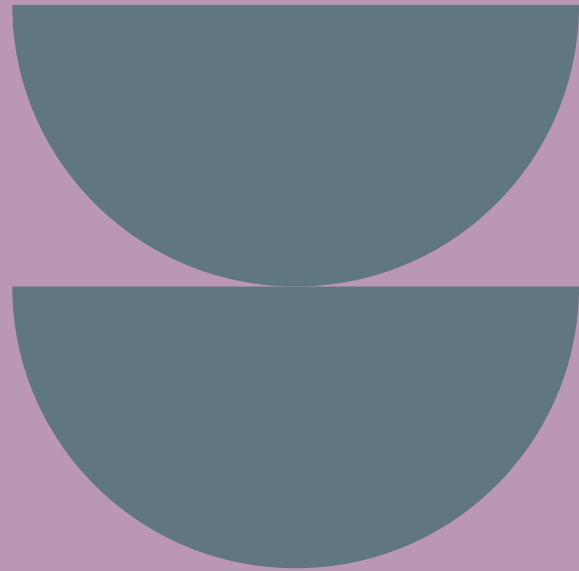
In the past 18 months, advancements in generative AI technology have drastically lowered the cost of transcription, translation, and automated lip-syncing. Although many of these tools are marketed as multilingual, our testing has indicated that they require significant customisation to accurately capture the nuances of Welsh pronunciation.

What if these tools could be refined specifically for Welsh-to-English and English-to-Welsh translation and made available as a specialised platform for translation agencies and dubbing artists? This capability could provide clients with higher-quality services, addressing a unique market need.

We believe that refining AI lip-syncing technology for Welsh translation not only offers potential for market expansion but also contributes to making content more accessible and inclusive for a broader audience.

Opportunity in the TV and Media Landscape

There is substantial potential in adapting AI-driven lip-syncing and dubbing technologies to enhance the international distribution of television productions originally created in Welsh. This project aims to harness that opportunity, promoting cultural exchange and expanding the reach of Welsh media worldwide.



Exploration

The Llais project focused on identifying potential users, exploring viable business models, and assessing the value of high-quality AI-assisted lip-sync and dubbing technology for Welsh-language content.

Research activities included desk-based market analysis, competitor research, face-to-face interviews, and the creation of sample outputs to benchmark the capabilities of existing AI tools.

Key Research Questions

The project sought to answer several key questions to gauge the relevance, value, and accessibility of this service across different markets:

- **Audience Benefits:** What are the potential advantages for end-users, and how can these be effectively communicated?
- **Market Standards:** Can we deliver these advantages at a price and quality level that could set new expectations for lip-synced content?
- **Ethical Considerations:** How can we employ AI ethically within this service?
- **Business Model Viability:** What business model would be most appropriate?




Benefits for the Final Audience

Initial research evaluated current AI tools for Welsh, focusing on their suitability for improving accessibility. The project considered alternative approaches to conventional practices, such as the use of embedded subtitles in video, which often misapply accessibility features meant for the hearing-impaired. The Llais project aims to replace these practices with accessible, high-quality lip-synced content, benefiting both social and cultural representation for Welsh-language media worldwide.

Market Exploration

To identify potential clients and stakeholders, interviews were conducted with a variety of Welsh TV production companies and notable figures within the media industry.



Findings by Industry Stakeholder

Television Production Companies

Research participants included a broad spectrum of production companies varying in size and type, all involved in Welsh-language content.

The spectrum included the back to back filming of drama content, entertainment programmes, animation, live programming, shortform content, corporate and education videos.

Budgets are being cut across the sector and companies are thinking creatively about ways to save money in the production process with compromising quality.

Companies expressed interest in the potential of the Llais project to reduce production costs and time, with an eagerness to explore practical applications.

There was a general consensus that the project would have an impact on both cast and crew. There was interest as to how the project would be received amongst the wider sector.

There was no immediate concern about costs; instead, the focus was on the platform's potential to broaden content accessibility and appeal to a wider audience.

However, stakeholders cautioned that the product should only be launched when it is market-ready to avoid premature entry.

Stakeholders also shared ideas for the platform's use in educational videos, Welsh learner-targeted content, factual programming, and international distribution.

Questions were asked if the project could be used in the animation genre and also would Llais eventually be available for Live programming.

Llais isn't about the use of AI Voices or cloning voices which was a common misconception.



Content Genre Impact

Drama

Some Welsh productions, like drama series, are filmed in both Welsh and English. Implementing Llais could streamline this process, reducing costs and the time commitment for cast and crew. Unions within the creative sector noted a potential impact on job roles, underscoring the need for industry engagement.

Factual Programming

Factual programs, which involve minimal cast interaction, often use voiceovers recorded in studios. Llais could significantly improve viewer experience by enabling seamless dubbing.

Corporate Training Videos

As an essential income source for many producers, corporate training videos could benefit from Llais, enabling multilingual production. Currently, subtitles are often added post-production, which can compromise accessibility. Llais would enable natural language dubbing, improving accessibility and viewing experience.

Educational Content

Similar to corporate training videos, educational videos could benefit from Llais for reversioning into Welsh. Some production companies suggested repurposing archived content for Welsh learners, potentially unlocking new avenues for commissioning.

Individual Perspectives

Actors and Dubbing Artists

Voice actors and dubbing artists expressed mixed feelings about AI.

Voice artists they have been aware that AI developments have been on the horizon for the past 2 years.

While initially wary, some actors see AI as complementary, with opportunities to apply their skills in areas like audiobooks, which remain less impacted by AI.

There is resistance from actors' unions toward AI-driven approaches, highlighting a need for industry education to help actors navigate this shift.

There is currently a number of AI voice cloning jobs for voice actors. Many actors won't take the work but others are getting in early.

The number of voices available in Welsh is minimal therefore dubbing work in the Welsh language hasn't yet been affected.

Translators and Writers

In the translation sector, AI tools like Google and Bing Translate are increasingly common but require human refinement.

Writers are concerned about AI's role in downgrading Welsh in translation work, as well as the use of scripts for AI training without sufficient input and permission from writer and translators.

Timing and synchronisation are particularly important in animation,

where writers collaborate closely with producers to ensure accurate lip-syncing.

As an actor and writer in Wales there is a requirement to undertake lots of different work - voice over, translate, write, act, radio, tv, theatre etc - in order to make a living.

Broadcasters

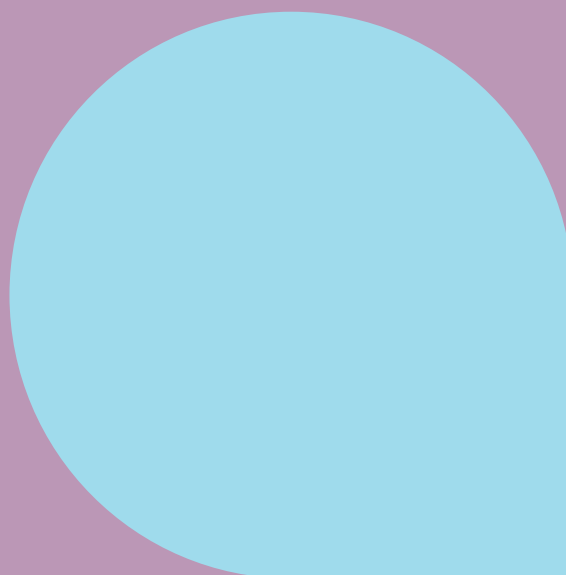
A senior member of the management team at BBC Cymru Wales oversees all aspects of artificial intelligence (AI) integration within the organisation. Currently, there is limited use of AI, but various research and innovation projects are underway to explore its applications in administrative functions. The primary focus of these initiatives is to enhance efficiency in administrative roles, thereby allowing for increased investment in production and creative areas.

However, there exists a sensitivity regarding the impact of AI on the

creative industries, particularly within the BBC. The organisation collaborates closely with relevant unions to address these concerns. Additionally, there is a cautious approach towards the sources of data utilised for AI training. Projects involving AI in the creation of metadata for the BBC archive have raised significant concerns, prompting the organisation to work with partners to ensure that the archive data is not misused or trained beyond the organisation's parameters.

S4C Engagement

Regrettably, attempts to engage with a representative from S4C regarding the project were unsuccessful, as there was no response to the requests for discussion.



Welsh Government

The Welsh Government engaged in discussions regarding the Llais project with the 2050 project team, emphasising the importance of Welsh language technology. Specialists in Welsh language technology have been informed about Llais and will closely monitor its development.

To support AI advancements that accommodate the Welsh language, the Welsh Government has collaborated with

OpenAI to enhance language processing capabilities in the GPT-4 chatbot.

Additionally, the Welsh Government is actively developing a policy framework on AI that prioritises ethical standards, responsible AI deployment, and social partnership. This policy approach includes a commitment to working alongside trade union partners to responsibly integrate AI tools and services.

Unions

Conversations were held with representatives from several unions representing the creative sector to gauge their perspectives on the integration of AI. The Llais project seeks to develop a business model that considers the ethical and moral implications of employing AI lip-syncing tools, while also aligning with existing policies and best practices for future implementation.

Representatives from BECTU (Broadcasting, Entertainment, Cinematograph and Theatrical Union), the TUC (Trades Union Congress), the Writers Guild of Wales and the

UK, and the Welsh language umbrella organisation Teledwyr Annibynol Cymru (TAC) provided valuable insights into their members' views. All representatives expressed interest in the Llais project and its proposed approach.

While the unions aim for constructive engagement, there are notable concerns regarding the use of AI in creative processes. It was acknowledged that achieving a solution that satisfies all stakeholders may be challenging.

Key Concerns Raised by Union Representatives

Compensation: Ensuring fair remuneration for all individuals involved in the production process is paramount. There are concerns that the shift towards AI could result in job losses not only for cast and crew but for all contributors to the creative process. While financial statements may appear stable, the expenditure may shift towards the tech industry rather than supporting the creative sector.

Training Data: There are significant apprehensions regarding the origins of training data and its copyright implications. Reports from union members indicate instances of data scraping without compensation, leading to confusion regarding copyright ownership and credit attribution. It is crucial to scrutinise the sources of training data to ensure it is ethically sourced and properly compensated.

Misuse of Data: Ethical considerations surrounding data misuse are also critical. There must be assurances that data is utilised in accordance with contractual agreements and within specified limitations, preventing it from being sold or repurposed elsewhere.

Transparency and Credit Attribution: Union representatives emphasised the importance of transparency in AI usage. Proper crediting of individuals' contributions and clarity regarding how AI has been utilised in productions is essential to foster trust among audiences.

Education and Awareness: There is a pressing need to educate individuals in the creative sector about AI technologies, particularly those who will be most impacted yet remain unaware of the implications. BECTU is exploring engagement strategies with its members to develop guiding principles and a comprehensive policy on AI usage. The Writers Guild and Equity have already established policy statements regarding AI, while TAC is collaborating with the Writers Guild in Wales to create principles for their members.



Testing AI Models

In light of our discussions with various stakeholders, we aimed to establish a robust working model for the Llais project. The insights gleaned from our interactions with union representatives and industry professionals in Wales informed our development of contracts for the actors involved in the Llais filming process. These contracts explicitly outlined the following:

- **The data training processes associated with the project.**
- **The intended platforms for video distribution.**
- **Our commitment to transparency regarding the use of artificial intelligence (AI).**

It is important to note that all videos produced will be labeled as research and development.

The Llais model emphasises the use of the same actor's voice in both Welsh and English versions, explicitly excluding the use of AI-generated voices and ensuring the ethical handling of all data.



Testing Process

For the testing phase, Llais filmed two actors: a female and a bearded male, both proficient in Welsh. The filming took place in an innovation studio in Cardiff, with a professional camera operator overseeing the process.

The actors were recorded in both languages to facilitate effective comparison and assessment. Four different AI models were employed during this testing phase to evaluate their effectiveness and performance.

Testing Video Examples

A comprehensive testing report detailing the findings and lessons learned from this phase can be found in the testing tech report.



News Comparison Reveal

<https://youtu.be/PLOQvcaBvIU>



AI Model 3 - Actor A - Walk

<https://youtu.be/6176pKS1FQY>



Actor B - Walking

https://youtu.be/p9osgq_rV7g



Actor B - News

<https://youtu.be/GoGi1mp2MMI>



Conversation

<https://youtu.be/bmIYzN8HId4>



AI Comparison - Walking

<https://youtu.be/rzR4O03Tz9A>



AI Comparison – News

<https://youtu.be/XhWo-7rLiaY>



AI Comparison – Facial Hair

<https://youtu.be/kPPj7BE6GVs>

The Llais project has yielded several important insights from both the research conducted and the testing process. These learnings will play a crucial role in shaping the project's development and informing the creation of a viable business model.

- 1. Further Testing Needed:** Additional testing with a broader range of video content is necessary before the project can engage with the market effectively. This will help ensure that the AI models are robust and reliable across diverse contexts.
- 2. Collaboration with Industry Professionals:** The success of the Llais model hinges on close collaboration with translators and dubbing artists. Engaging these professionals will be essential to guarantee the availability of high-quality data for training AI systems, thereby enhancing the overall effectiveness of the lip-syncing process.

By addressing these key areas, the Llais project is positioned to refine its approach and establish a strong foundation for future market engagement.

Credits:

Actor: Tom Mumford

Actor: Melanie Owen

Film: Ian Smith, Auntie Margaret productions

Tech partner: Robin Moore, Shwsh

Project mgt: Aled Jones, Y Pod Cyf

Project Summary

The Llais project's research illustrates the significant potential for AI-driven lip-syncing in expanding the reach and impact of Welsh-language media. With ethical considerations, accessibility improvements, and a market-ready product, Llais could reshape how Welsh content is experienced locally and internationally, making it more inclusive and widely accessible across different media sectors.

The Llais project has garnered interest due to its potential positive impact, although there remains an awareness of its implications for the creative sector. There is a consensus that Welsh language initiatives should lead in the adoption of technological advancements within the creative industries, and individuals within the sector intends to monitor the development of the Llais project closely.

If we were to do it again, how would we approach it differently?

Given another opportunity, we would extend the project timeline. This would allow us to engage with key stakeholders, such as S4C and Equity, which we couldn't coordinate with during the original timeframe but plan to connect with afterward.

A longer timeline would also give stakeholders more time to provide feedback on the examples developed through Llais.

Our engagement with stakeholders served as an initial market exploration to gauge interest in the concept.

We are excited about the future of the Llais project.

The examples created have exceeded expectations and are poised to spark conversations within Wales's creative media sector.

Further Reading and Resources

Equity: Equity's AI Vision Statement

<https://www.equity.org.uk/advance-and-support/know-your-rights/ai-toolkit/equitys-ai-vision-statement>

Writers Guild: Writers and AI report

<https://writersguild.org.uk/ai/>

Writers Guild: Writers and AI – A policy statement

https://writersguild.org.uk/wggb_campaigns/writers-and-ai/

UK Government: A pro-innovation approach to AI regulation

<https://assets.publishing.service.gov.uk/media/64cb71a547915a00142a91c4/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf>

Pact Census 2024: AI Annex

This is the AI Annex to the 2024 Pact TV Production Census, a report by Oliver and Ohlbaum Associates.

<https://www.pact.co.uk/assets/03235A5D%2D7494%2D411D%2DA-CA23A967EA43C1E/>

Welsh Government: Welsh language better prepared for artificial intelligence developments, thanks to Welsh Government investment

<https://www.gov.wales/welsh-language-better-prepared-artificial-intelligence-developments-thanks-welsh-government>

Welsh Government: ChatGPT learns Welsh

<https://www.gov.wales/chatgpt-learns-welsh>



Appendix A - Technical Testing Lessons

Comparison of 4 different AI tools all claiming to be state-of-the-art. Initial research include a wider array of tools, but these were discounted. We are not including a number of the best-known tools (e.g. Synthesia, RASK) as they only allow the use of an AI voice, a route we don't wish to go down due to poor quality for audio performance and impact on jobs. We did some initial test with tools which animate images (e.g.

ReTalker, SadTalker) to audio and this may have a usecase, but didn't not fit with doing a live action shoot or testing a service aimed at TV.

Trials specifically dealt with difficult scenarios with different gender, ethnicity, facial hair, moving camera, faces in profile (side view), multiple people in shot and varying lighting.



Tool	Price	Pros	Cons
Model 1	\$49/month for 65 minutes	<ul style="list-style-type: none"> • Lip accuracy reasonably good • Quick processing • Usable on more distant faces 	<ul style="list-style-type: none"> • Poor mask on larger faces and when moving • Limited when face is in profile • Jittering on beard and large faces • Mouth can look a little cartoony • No multiple people
Model 2	\$28/month for 7.5 minutes	<ul style="list-style-type: none"> • Lip accuracy reasonably good • Quick processing • Can handle multiple faces (max 5) • Usable on more distant faces even when moving 	<ul style="list-style-type: none"> • Fails on larger videos • Fails if faces are in profile • 1080p output only • Mouth noticeably over sharpened and high contrast, while mask can appear blurry
Model 3	\$24 per actor/scene + \$8 per minute	<ul style="list-style-type: none"> • Excellent lip accuracy in most instances • Good with close up and when moving • Much wider masking, almost flawless mask • Good Welsh adherence 	<ul style="list-style-type: none"> • More complex workflow • Limited when face in full profile
Model 4	Free	<ul style="list-style-type: none"> • Local opensource model - Free • Reasonably accuracy mouth shape on English • Extensible – can swap out or retrain AI tools 	<ul style="list-style-type: none"> • No Welsh • Very poor mask • Blurred on beards • No profile faces • No multiple people

Model 3 is clearly the leader using a mixture of lip-sync and deepfake AI to do the masking. It could get very expensive with the wrong workflow, but it delivered very good results in most cases even without much footage for the per actor training. Potentially a better voice recognition model for Welsh would help, or at the very least a tailored approach to capturing training data per actor to ensure all Welsh letter sounds are captured.

Ideally for corporate market where quality expectation and the budget available may be lower the other models could be used in specific instances.

Model 1 and model 2 could be usable with more distant moving faces and offer an alternative for front facing video with no beard.

Model 4 is the state-of-the-art opensource model and would be worth working on further in longer-term to deliver a low overhead basic model. It is the most extensible e.g. while it currently using a model for interpreting the audio which cannot handle non-english languages at all, it could easily be rewritten to incorporate a different audio language model.

Lots of lessons for how to do this and run the shoot, to ensure we can deliver high quality lip sync when dubbing:

- We need a minute of outtakes for training for each scene/set-up or just the performers talking before each shot (in the position/lighting they will be filmed in). For front facing single person shots trained wasn't always necessary, particularly with more natural lighting.
- Use .wav rather than .mp3 to get best audio results, and ideally when on set get a clean audio feed – better for the AI and easier for any editing of the audio to match the timing.
- We need to get the dubbing artist to record the track for each actor on a separate audio file and the better the timing the better the result. Aligning the main 'beats' of the dialogue in both languages was key in making the lip sync look natural, and this can be very time consuming.
- Welsh and English tend to take quite different amounts of time to say the same thing. While Welsh can be quicker, it generally it took ~10% longer, this is enough to make it difficult to squeeze all the Welsh audio into the English video, or for English to Welsh leaving awkward pauses. While speeding up / slowing down audio of video and extending breaths is possible for corporate content, it won't meet the quality necessary for TV. We need to instead look at a pipeline for rewriting the dubbing script before recording and then recording translations to time, at a similar level of accuracy used in ADR sessions.
- Otherwise matching the speed of the audio and length of breaths to get the key 'beats' of the movements in line with the dubbing audio, takes a lot of time, and needs to be considered as a key cost. In the tests there are instances of the lips looking wrong, because the lips moved unnaturally fast.
- The length of audio files needs to be the same as the video - it may be easier to edit together – and I ended up exporting audio from Premiere.
- All in all for a professional result, ideally we should be planning to do this language revisioning before the programme is filmed and ensuing extra material is captured.

Next Steps

- Ideally transcription and translation should be part of the pipeline, and we need to do further development on an end-to-end pipeline for matching timing and potentially rewriting scripts (with a human writer/translator), which then flows into Model 3.
- For lower budget/quality projects we should do more work on a pipeline using models 2 & 4 specifically focusing on the problems of getting the Welsh language supported properly.
- We now know this is feasible and can deliver the necessary quality, but still need to know if it is financially and operationally viable in practice and for which scenarios is it valuable. Further testing on a more complex programme would help price up the wider service and explore other issues and optimisations that will come up in real-world situations.
- The scenario we testing had all the audio captured up front alongside the video. We need to test how this would work with an ADR style session for capturing audio after the shoot.
- We need to explore the models with real broadcast content, short form extra programme content and corporate videos.
- Explore funding options, investment and partnerships to streamline the model to create a viably commercial option.

