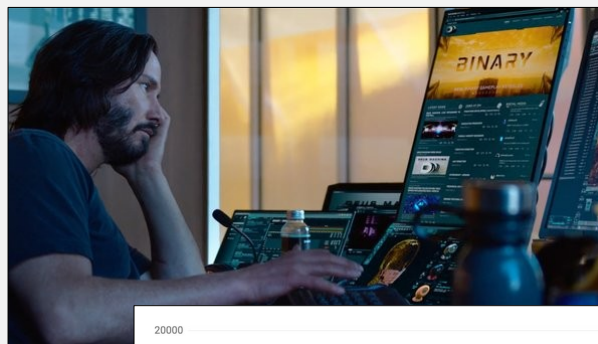


# Introducción

# Introducción

- Contexto
- Motivación
- Objetivos



## Restrictions on media

Suspension of broadcasting in the EU of state-owned propaganda outlets:

- Sputnik
- Russia Today
- RTR Planeta
- Russia 24
- TV Centre International



# Introducción

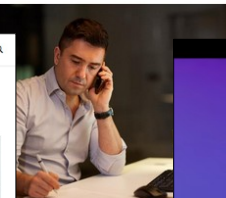
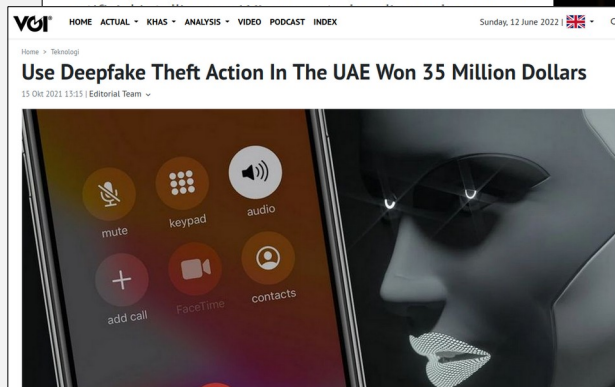
- Contexto
- **Motivación**
- Objetivos

## Unusual CEO Fraud via Deepfake Audio Steals US\$243,000 From UK Company

September 05, 2019

An unusual case of CEO fraud used a deepfake audio, an

Related Pos



# Introducción

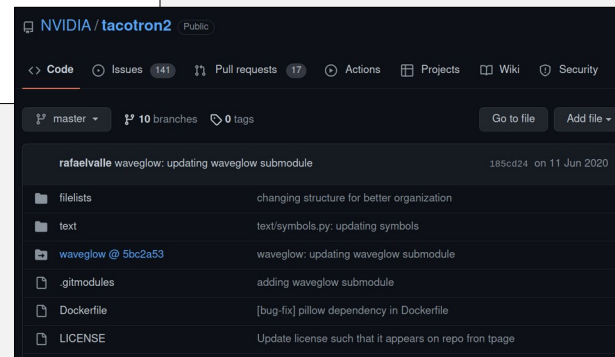
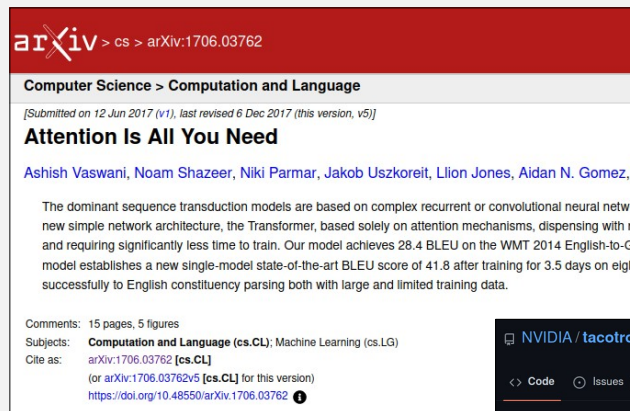
- Contexto
- Motivación
- **Objetivos**

- Explorar el Estado del Arte en síntesis de voz
  - Revisión histórica
  - Última investigación
- Explorar las necesidades particulares para trabajar en castellano
  - Adaptación de modelos
- Probar los modelos y herramientas existentes
  - Reproducción de resultados
  - Aportaciones, cambios y mejoras

# Metodología

# Metodología

- Investigación
- Experimentación





# Estado del Arte



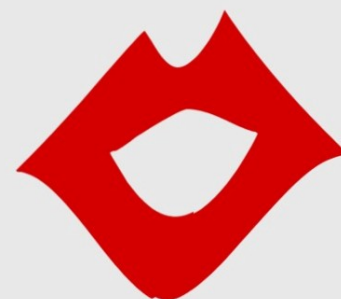
# Estado del Arte

- **Modelos no entrenados**
- Modelos entrenados
  - Por componentes
  - Extremo a extremo

## Welcome to MBROLA project

### Table of contents

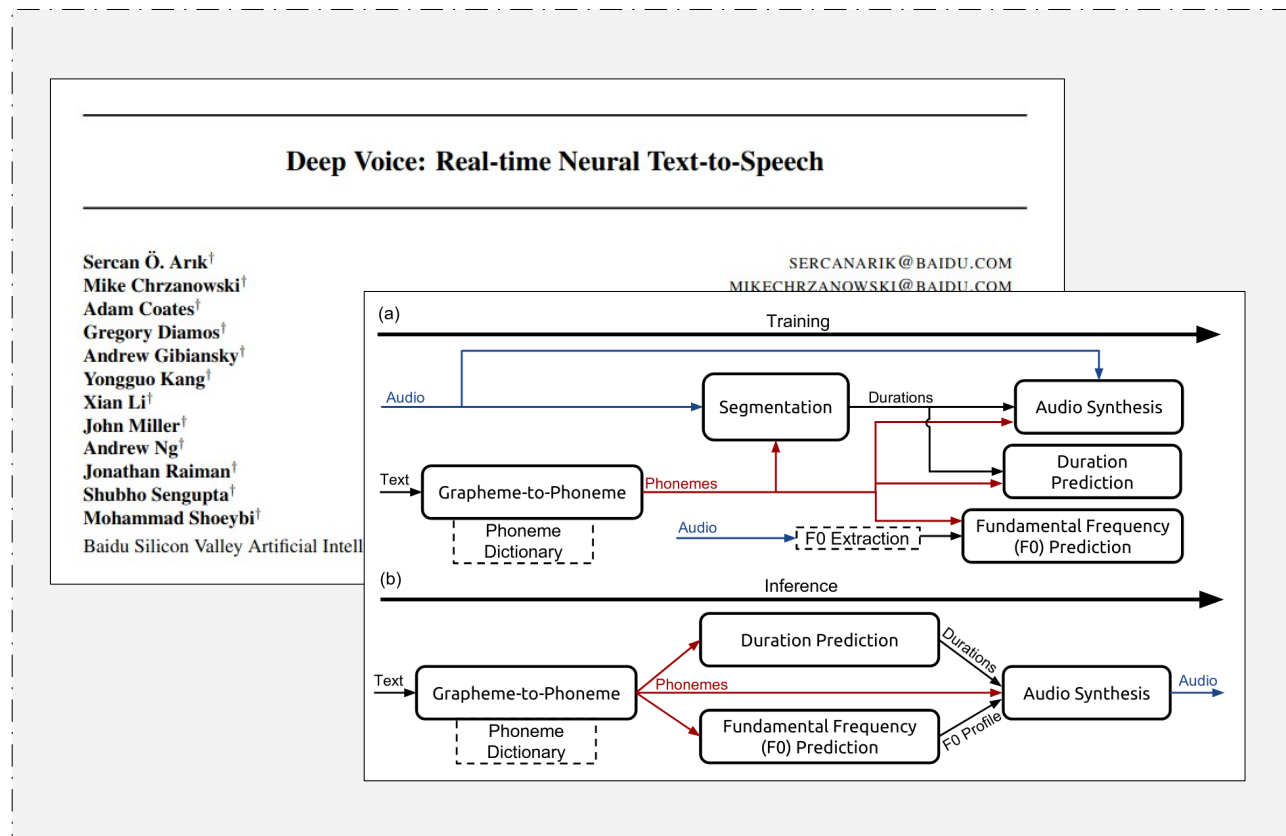
- A brief description of MBROLA
- Compilation
- Installation
- Execution
  - Using Pipes
  - Renaming and Cloning phonemes
  - On MSDOS/Windows
  - On modern Linux



**eSpeak**

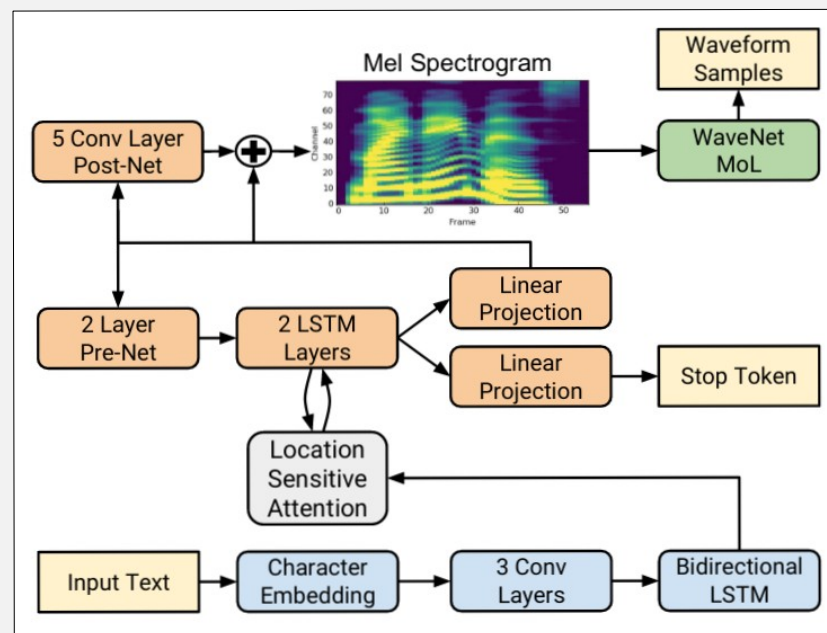
# Estado del Arte

- Modelos no entrenados
- Modelos entrenados
  - **Por componentes**
  - Extremo a extremo



# Estado del Arte

- Modelos no entrenados
- Modelos entrenados
  - **Por componentes**
  - Extremo a extremo



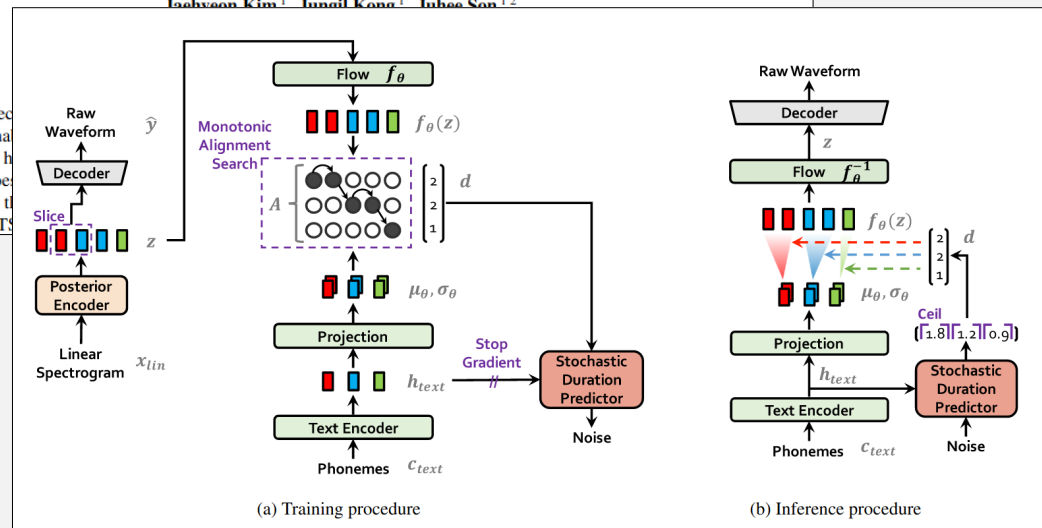
# Estado del Arte

- Modelos no entrenados
- Modelos entrenados
  - Por componentes
  - **Extremo a extremo**

## Conditional Variational Autoencoder with Adversarial Learning for End-to-End Text-to-Speech

Jaehyeon Kim<sup>1</sup>, Junil Koo<sup>1</sup>, Juhoo Son<sup>1,2</sup>

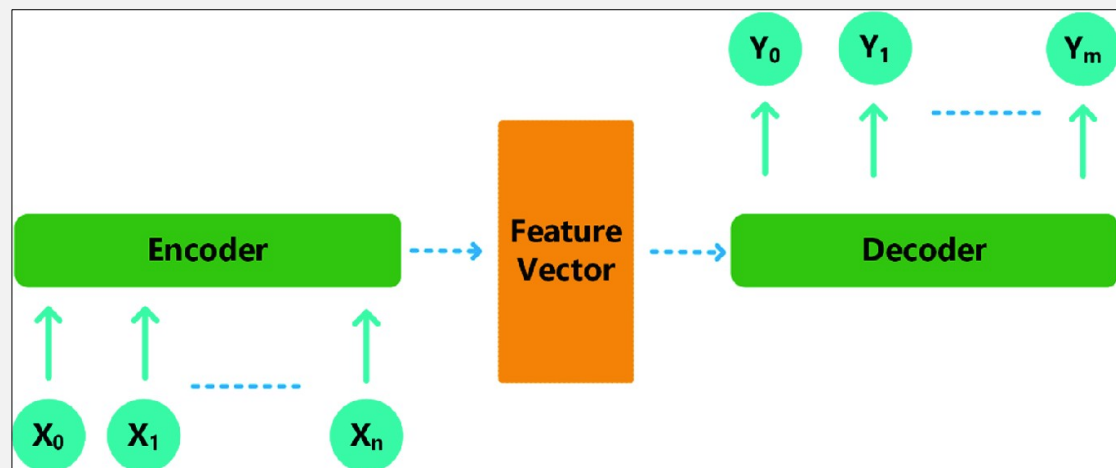
Several recent models enable sampling high quality waveforms. In this work, we propose an end-to-end TTS model.



# Resultados

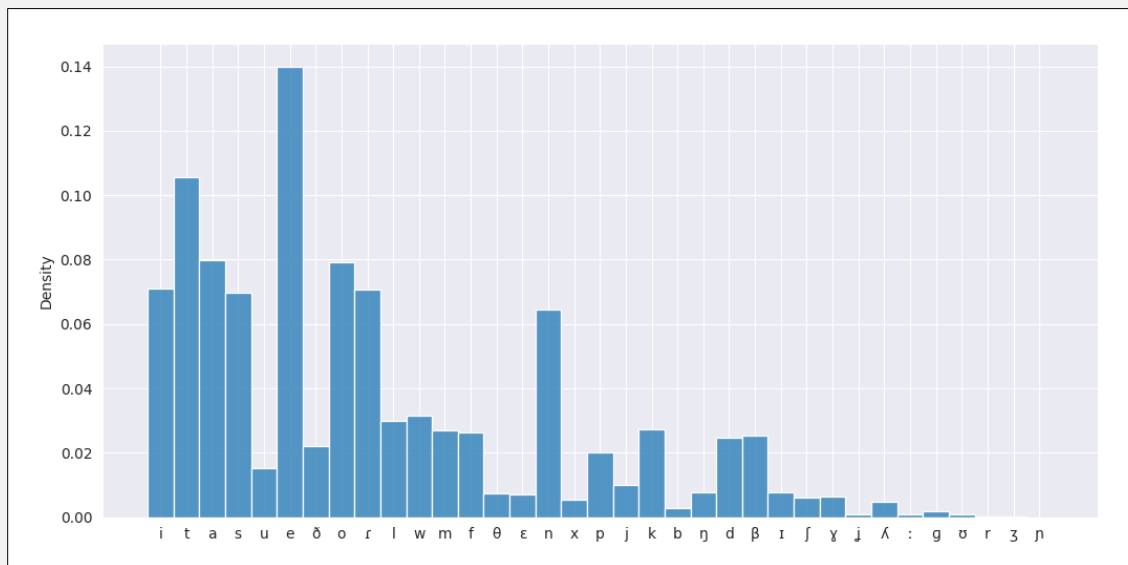
# Resultados

- Resultados de investigación
  - **Arquitectura E/D**
  - Datasets
  - Herramientas comerciales
- Resultados experimentales
  - Datasets
  - Modelos



# Resultados

- Resultados de investigación
  - Arquitectura E/D
  - **Datasets**
  - Herramientas comerciales
- Resultados experimentales
  - Datasets
  - Modelos



# Resultados

- Resultados de investigación
  - Arquitectura E/D
  - Datasets
  - **Herramientas comerciales**
- Resultados experimentales
  - Datasets
  - Modelos



## Generate AI Voices that sound real.

Resemble AI supercharges your AI voice with a text-to-speech generator paired with real-time APIs to build immersive experiences.

**FakeYou**

Clone My Voice!

Create

Comunidad

Desarrollador



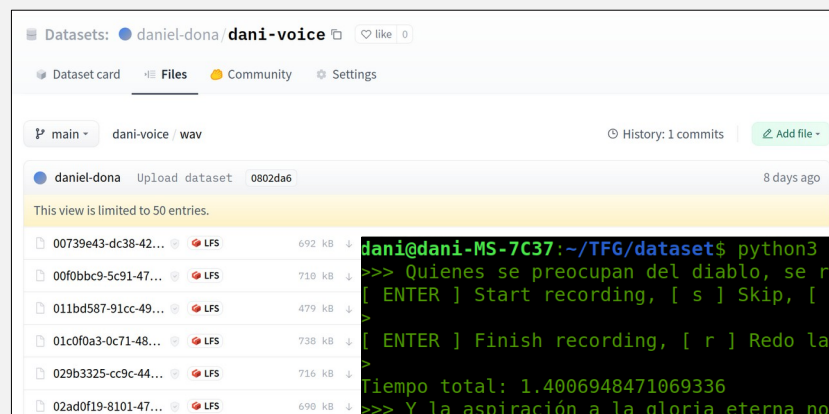
### Texto a Voz

Usa FakeYou tecnología falsa profunda para crear voces realistas con tus personajes favoritos



# Resultados

- Resultados de investigación
  - Arquitectura E/D
  - Datasets
  - Herramientas comerciales
- Resultados experimentales
  - Datasets**
  - Modelos



```
dani@dani-MS-7C37:~/TFG/dataset$ python3 simple_record.py
>>> Quienes se preocupan del diablo, se rien del microbio
[ ENTER ] Start recording, [ s ] Skip, [ q ] Quit
[ ENTER ] Finish recording, [ r ] Redo last sentence, [ q ] Quit
Tiempo total: 1.4006948471069336
>>> Y la aspiración a la gloria eterna nos aleja de la gloria humana
[ ENTER ] Start recording, [ s ] Skip, [ q ] Quit
[ ENTER ] Finish recording, [ r ] Redo last sentence, [ q ] Quit
Tiempo total: 2.8459529876708984
>>> Como el patriotismo, la pasión de la gloria debe sugerirse y nunc
[ ENTER ] Start recording, [ s ] Skip, [ q ] Quit
> s
>>> Mas la vida cenobítica resultaría para la mayoría de los sabios i
[ ENTER ] Start recording, [ s ] Skip, [ q ] Quit
█
```

# Resultados

- Resultados de investigación
  - Arquitectura E/D
  - Datasets
  - Herramientas comerciales
- Resultados experimentales
  - Datasets
  - **Modelos**
    - Reproducción de resultados
    - Ajustes

- Tacotron 2 (implementación NVIDIA)
  - Problemas de alineamiento
  - Calidad no muy buena
- VITS
  - Calidad bastante buena
  - Síntesis en tiempo real

# Resultados

- Resultados de investigación
  - Arquitectura E/D
  - Datasets
  - Herramientas comerciales
- Resultados experimentales
  - Datasets
  - **Modelos**
    - Reproducción de resultados
    - Ajustes

- VITS
  - Bandas MEL del espectrograma
  - Uso de diferentes frecuencias de muestreo
  - Ajustes de batch size y learning rate
  - Uso de grafemas o fonemas como entrada

# Conclusions

# Conclusiones

- **Importancia del trabajo**
- Líneas futuras

- Confirmación de la posibilidad de generar falsificaciones
- Accesibilidad de herramientas para la generación
- Falta de mecanismos de protección y posible aumento del uso

# Conclusiones

- Importancia del trabajo
- **Líneas futuras**

- Generación
  - Mejor dataset en castellano
  - Modelos de transferencia de conocimiento
- Detección
  - Investigación en técnicas de identificación
  - Herramientas accesibles

# Demostración

# Muchas gracias