The BBC's 'Virtual Voice-over tool' ALTO: Technology for Video Translation

Susanne Weber

Language Technology Producer, BBC News Labs

In this presentation….

- Overview over the ALTO Pilot project
- Machine Translation and Computer Assisted Translation
- Text to Speech synthesis
- Users' experience with this technology
- Conclusions

Automated Language Tool

Production tool for the translation of News videos

Collaboration between
- News Labs
- World Service
- Global News

Go to   http://www.bbc.com/japanese/video_and_audio/today_in_video

And http://www.bbc.com/russian/video_and_audio/today_in_video

We experimented with 2 types of News Videos

- **Short clips** without original narrator track

- **News Packages** containing several voices

Typical Workflow for Video Translation

Translate Script

Record Voice-over tracks

Align Audio & Video

Edit Audio

Balance Audio Tracks

src 00:00:18:05

©BBC

Typical Workflow for Video Translation

Translate Script → MT
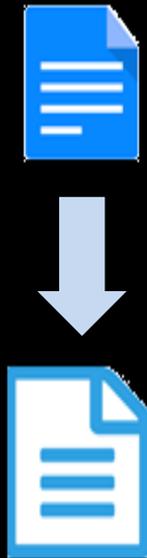
Record Voice-over tracks → TTS
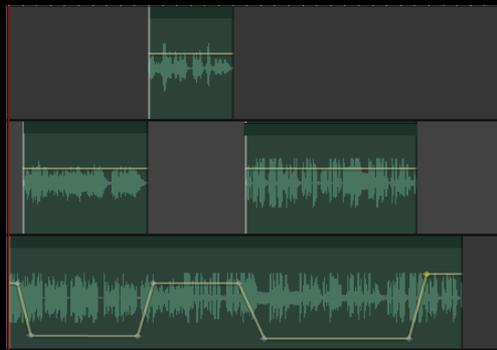
Edit Audio / **Balance Audio Tracks**

Align Audio & Video → FFmpeg

src 00:00:18:05
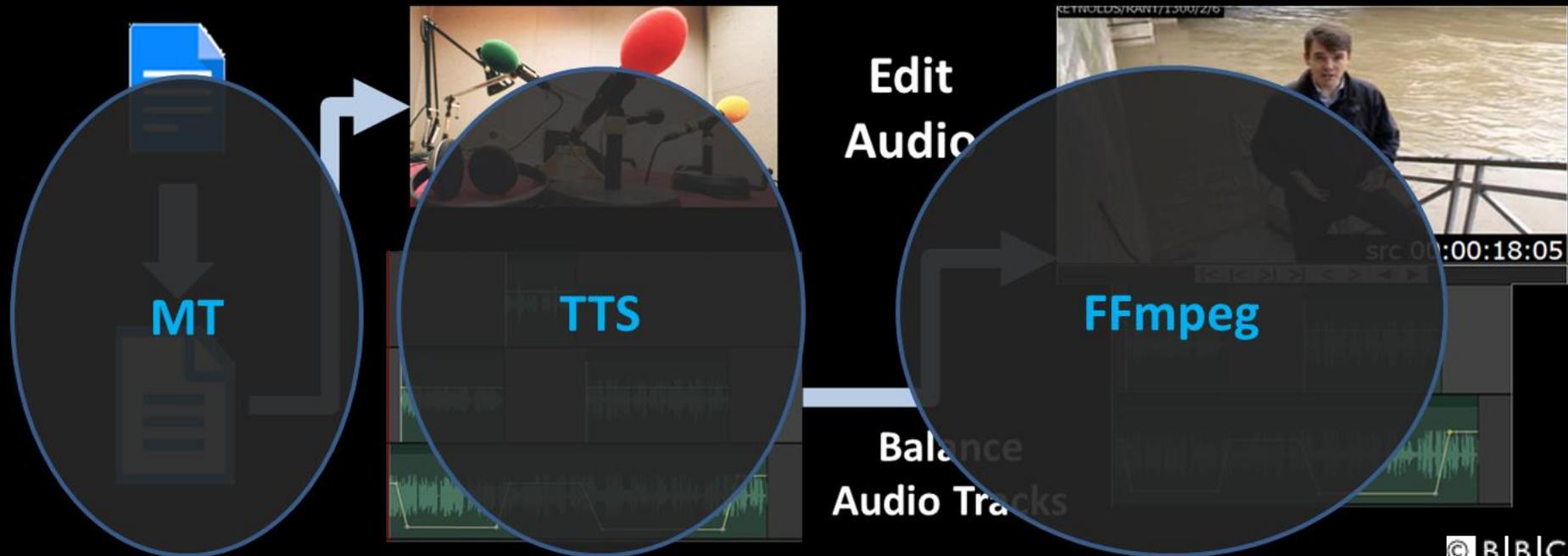
Off-the-shelf products

Text-To-Speech
Voice Synthesis

➢Google Translate

➢Computer-Assisted Translation

© B|B|C

Computer-Assisted Translation

Opium in Afghanistan Script

UK English en-GB

Opium production has long been a problem in Taliban-controlled areas of Afghanistan, but it is also grown in many government-controlled regions.

Some farmers even claim that the authorities have become more relaxed about the crops. These days they say who is in control of an area doesn't make that much difference to how much poppy is cultivated.

Afghanistan is expected to produce more opium - and therefore heroin this year than the world actually consumes.

+ Add new script line

Use RETURN to insert a new script line
Use SHIFT + RETURN to insert a line break
Use DELETE or BACKSPACE to remove an script line

© BBC

Russian ru

Play script

0 (1 seconds)   Silence
1 secs

1 (9.48 seconds)   Yuri

Производство опиума, - давняя проблема в подконтрольных талибану частях Афганистана. Но она есть ии в районах под контролем правительства.

10.5 (13.19 seconds)   Yuri

Некоторые фермеры даже утверждают, что подход властей в этом вопросе смягчился. Они говорят, что урожай мака не так уж и зависит от того, кто хозяйничает на иих территории.

23.7 (1 seconds)   Silence
1 secs

Video   Preview

0:35

Ducked Volume: (5 %)

Fade Down: (onset: 1s, total: 2s

Fade Up: (ending: 1s, total: 2s

🎤 Mixdown

Mixdown video clip:

Opium in Afghanistan Script

Computer-Assisted Translation

UK English en-GB

Opium production has long been a problem in Taliban-controlled areas of Afghanistan, but it is also grown in many government-controlled regions.

Some farmers even claim that the authorities have become more relaxed about the crops. These days they say who is in control of an area doesn't make that much difference to how much poppy is cultivated.

Afghanistan
heroin this y

How Good Is it???

+ Add new script line

Use RETURN to insert a new script line
Use SHIFT + RETURN to insert a line break
Use DELETE or BACKSPACE to remove an script line

© BBC

Russian ru   ◄) Play script

⏱ 0   (1 seconds)   👤 Silence ▼

         1 secs

⏱ 1   (9.48 seconds)   👤 Yuri ▼

         а в подконтрольных
         ть ии в районах под

⏱ 10.5   (13.19 seconds)   👤 Yuri ▼

Некоторые фермеры даже утверждают, что подход властей в этом вопросе смягчился. Они говорят, что урожай мака не так уж и зависит от того, кто хозяйничает на иих территории.

⏱ 23.7   (1 seconds)   👤 Silence ▼

         1 secs

🎥 Video   Preview

0:35

Ducked Volume: (5 %)

Fade Down: (onset: 1s, total: 2s

Fade Up: (ending: 1s, total: 2s

🎤 Mixdown

Mixdown video clip:

To put things into perspective…

- ca. 7,000 languages in the world

- Google Translate lists just over 100 languages

- Most TTS providers have fewer than 30 languages

Machine Translation – Computer Assisted Translation

High Resourced vs. Low Resourced Languages

- MT quality depends on:
    - Language Pairs
    - Source Text

Our editors' feedback:
- CAT is still faster than translating from scratch
- CAT is useful for proof-reading

Opium in Afghanistan Script

UK English en-GB

Opium production has long been a problem in Taliban-controlled areas of Afghanistan, but it is also grown in many government-controlled regions.

Some farmers even claim that the authorities have become more relaxed about the crops. These days they say who is in control of an area doesn't make that much difference to how much poppy is cultivated.

Afghanistan is expected to produce more opium - and therefore heroin this year than the world actually consumes.

+ Add new script line

Use RETURN to insert a new script line
Use SHIFT + RETURN to insert a line break
Use DELETE or BACKSPACE to remove an script line

© BBC

Russian ru

🔊 Play script

0 (1 seconds)   👤 Silence ▾

1 secs

1 (9.48 seconds)   👤 Yuri ▾

Производство опиума, - давняя проблема в подконтрольных талибану частях Афганистана. Но она есть ии в районах под контролем правительства.

10.5 (13.19 seconds)   👤 Yuri ▾

Некоторые фермеры даже утверждают, что подход властей в этом вопросе смягчился. Они говорят, что урожай мака не так уж и зависит от того, кто хозяйничает на иих территории.

23.7 (1 seconds)   👤 Silence ▾

1 secs

📹 Video   Preview

0:35

Ducked Volume: (5 %)

Fade Down: (onset: 1s, total: 2s)

Fade Up: (ending: 1s, total: 2s)

🎤 Mixdown

Mixdown video clip:

Japanese ja

🕐 0   (2 seconds)   👤 Silence ▾   ⬇   🔊

———□——————— 2 secs

🕐 2   (25.31 seconds)   👤 Kyoko ▾   ⬇   🔊

アメリカの国務省は、ヒラリー・   め、複数の
元国務長官が、メール利用にお   が不十分だ
ったとする調査結果を発表しま   ル使用記録
の問題点はクリントン氏の着任   点」だった
と書いています。

    Silence

    ✓ Kyoko (Female, ja)

    Ryo (Male, ja)

    Show (Male, ja)

    Takeru (Male, ja)

    Misaki (Female, ja)

🕐 27.3   (2 seconds)   ⬇   🔊

———□——————— 2 secs

➕ Add new script line

- It is difficult to get good quality voices – why is that?

- Currently, we are dependent on a small number of companies

- Why do some of them sound so natural, others don't?

- Why can't we have them in all the languages?

**There are 2 common methods for voices synthesis:**

1) Unit Selection

2) Statistical Parametric

Creating synthetic voices: Unit Selection

Record Voice

Pron Lexicon and word labels

Scripts (phonemes etc) to generate utterances data: "blah ... blah..."

Blah...blah

Utterance files
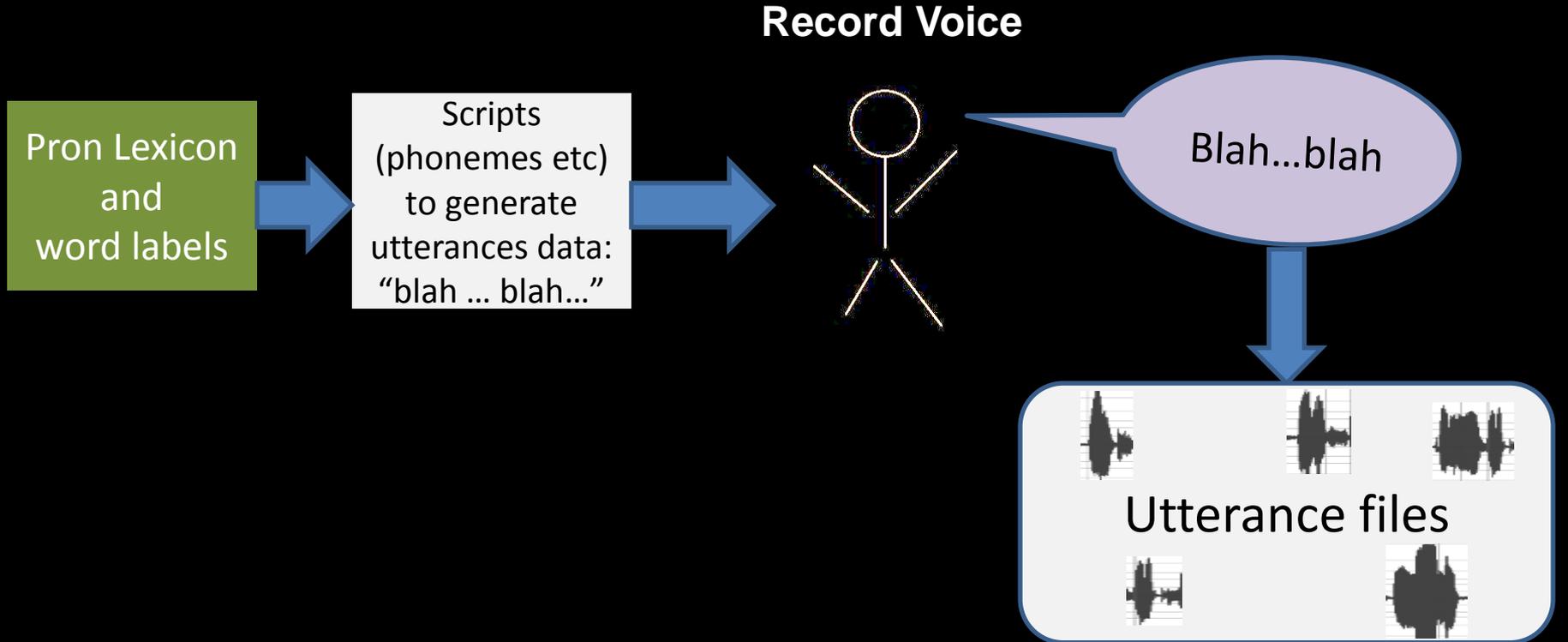
Text-To-Speech Synthesis: Unit Selection
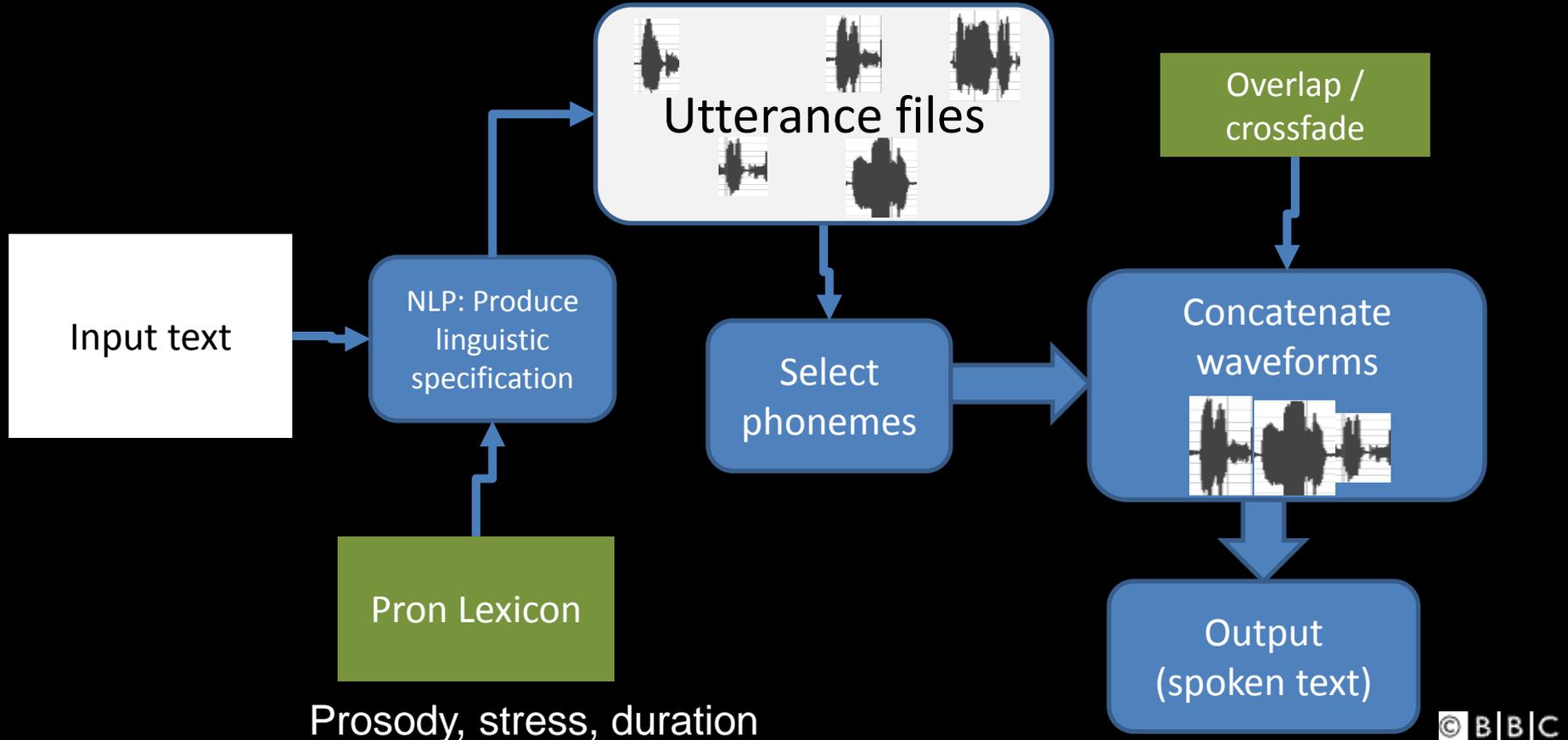
Unit Selection – Audio Examples

Japanese:

Unit Selection – User Feedback

- It sounds surprisingly natural

………*what is "natural"?*
There is no objective measurement
of "naturalness" – it is subjective

*……are accents "natural"? Scottish? Welsh?*
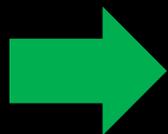
when they are human-like = "natural"

Unit Selection – Limitations

- TTS voices are emotionally neutral

- This is good for 'regular' news

- Unsuitable for emotionally charged contents, e.g. when voicing over victims of bomb attacks

- We have no control over their emotional expression in Unit Selection

Unit Selection – Phonetic performance control / Limitations

Spelling	Audio (English, UK)
Angela Merkel	🔊
Ang ella Markel	🔊
Vladimir Putin	🔊
Vladimeer Pootin	🔊
Francois Hollande	🔊
Francois O'Lond	🔊

©B|B|C

Training of Models: Statistical Parametric (simplified)

Speech Database

Speech Signal

Excitation Parameter Extraction

Spectral Parameter Extraction

Text / Words: LABELS

Training of TTS models

Hidden Markov Models

© B|B|C
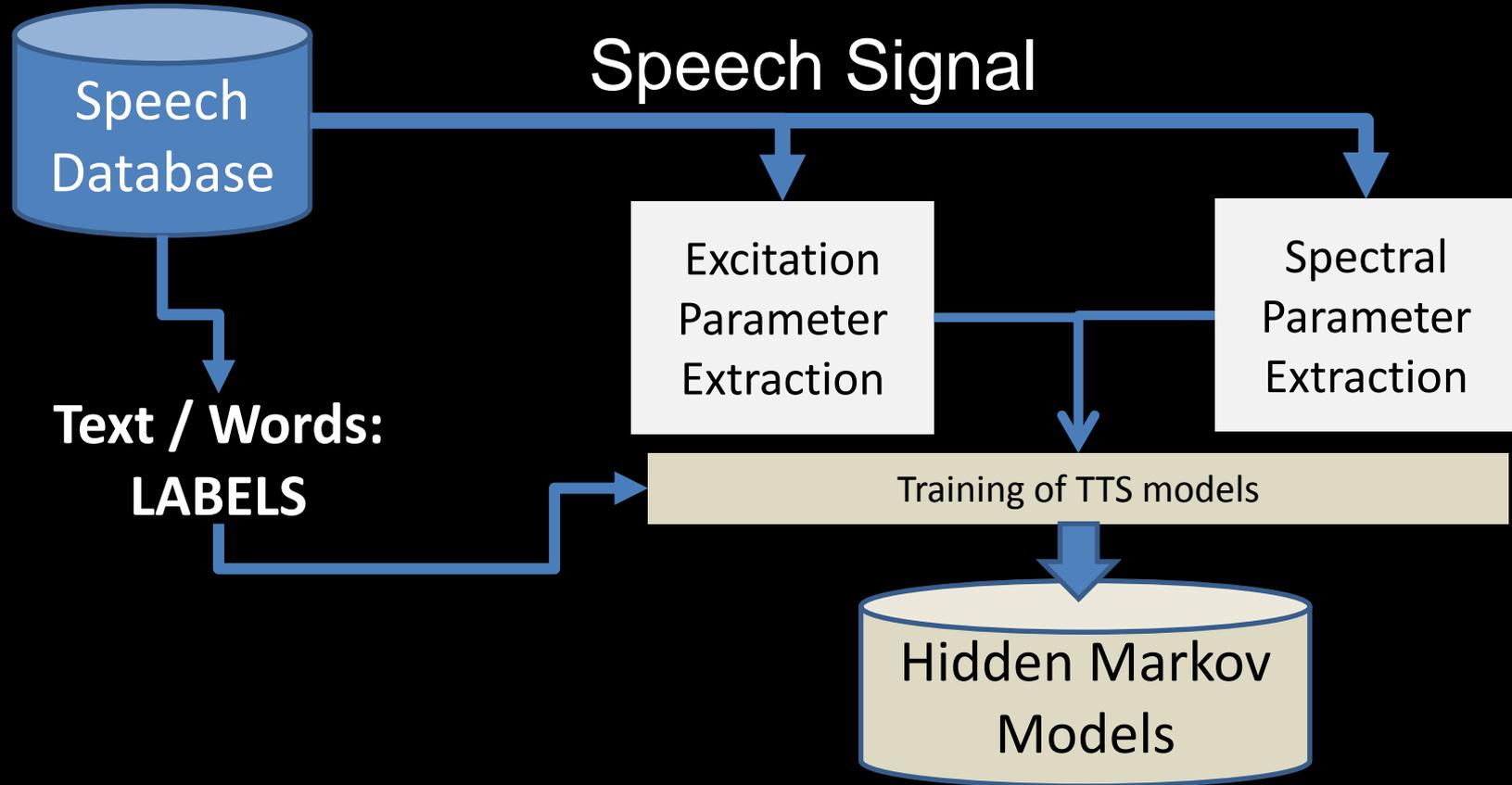
Voice Synthesis: Statistical Parametric (simplified)
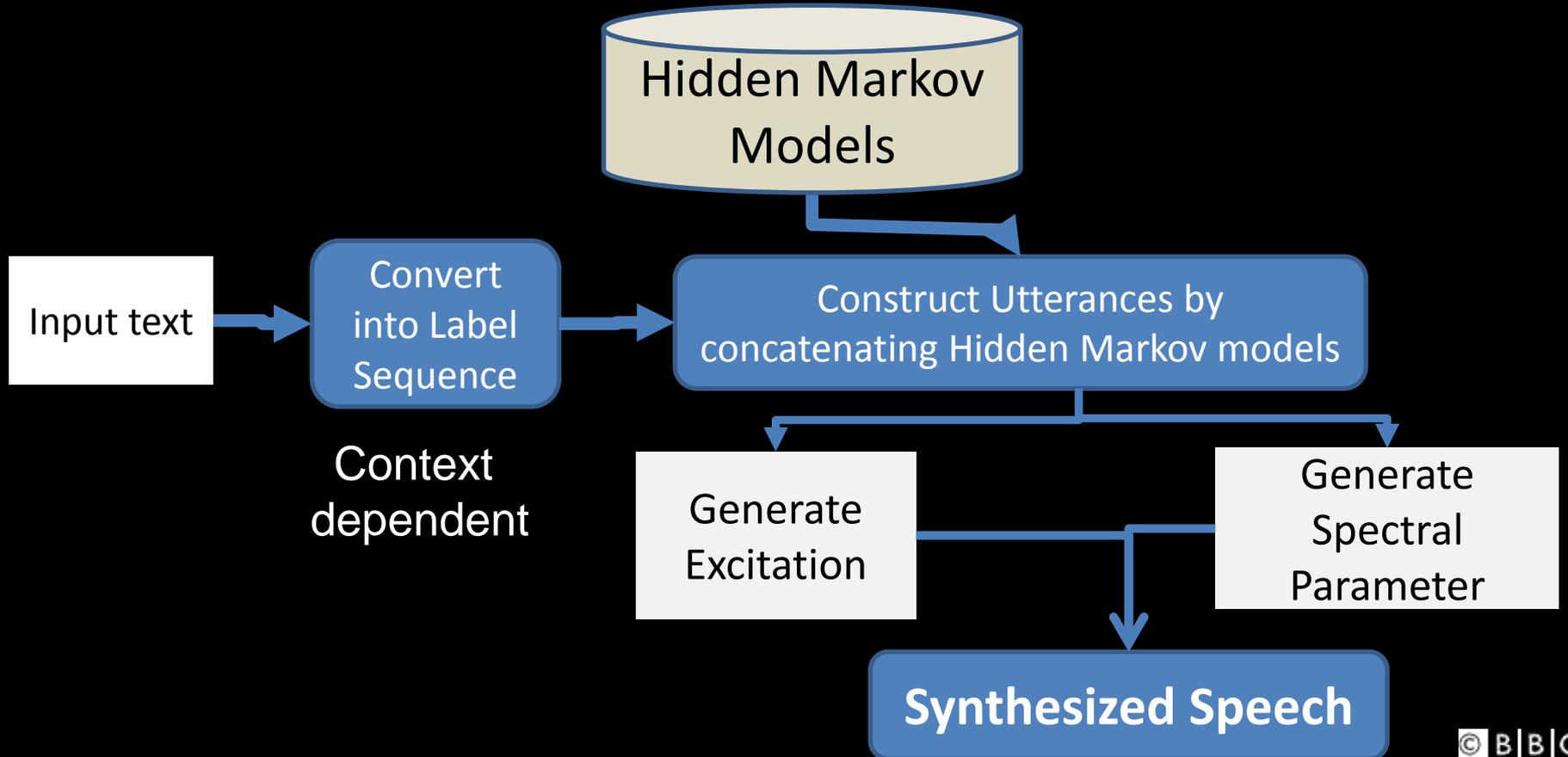
Statistical parametric TTS – the good bits

-   It is flexible, because of its statistical modelling process

-   It allows **expressive voices** to be generated;

-   the emotional expression of voices can be controlled

-   Voices are easier to build, because it doesn't need large amounts of datasets

-   this is good for low-resourced languages

Statistical parametric TTS – the sound

Audio examples:

Unit Selection	HMM
Japanese	Japanese
	Please go to this link:
	http://www.ai-j.jp/

Conclusion and Next Steps:

- We need language data for low resourced languages:
- For MT as well as TTS

- We need more languages and voices to be available

- We need expressive voices (e.g. a hybrid system)

- Collaborate with research groups and universities

- We want to tackle Graphics Translation
- And integrate automated transcription