Semi-automatic DVS Authoring Method

Inseon Jang, ChungHyun Ahn
Realistic Broadcasting Media Research Department, ETRI

jinsn@etri.re.kr

Younseon Jang
Department of Electronic Engineering
Chungnam National University

Outline

I*ntroduction*

B*ackground*

Procedure of the conventional DVS Production

P*roposed Method*

TTS based DVS

Non-Dialog Section Detection based on Audio/Subtitle Analysis

E*xample of the Implementation*

C*onclusion*

Introduction

- Several standardization activities to provide accessibility to broadcasting services for persons having disabilities
  - ITU-R BT.2207-2 (11/2012) Accessibility to broadcasting services for persons with disabilities.
  - http://www.itu.int/pub/R-REP-BT.2207-2-2012
  - http://www.itu.int/en/ITU-T/jca/ahf

(From ITU-G3ict Making Television Accessible Report)

Introduction

❖ Descriptive Video Service

- To provide visual media more accessible to those with seeing disabilities is to use audio descriptions which explain what is happening visually in the picture.

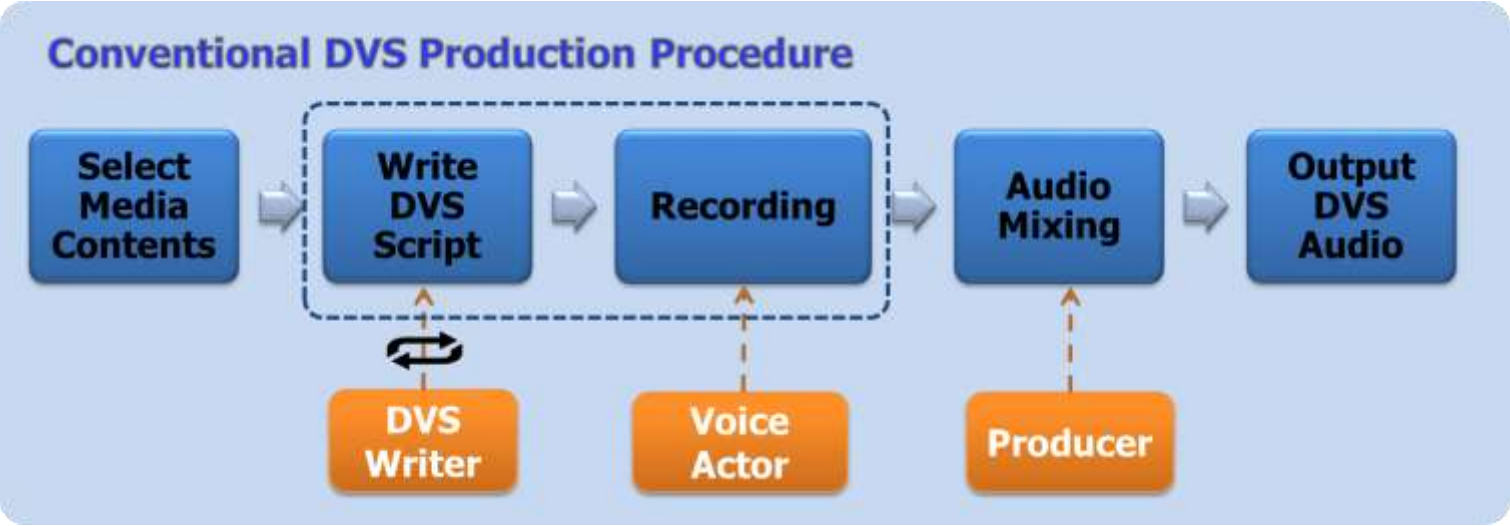
Provider		Target Provider	Start	Measure	Target of the final organization ratio(%)			Accomplishment
					Subtitles	DVS	Sign Language	
Terrestrial	Mandatory	Center	2012.1.	2012.7.	100	10	5	2013.12. (DVS: 2014.12.)
		Local	2012.1.	2012.7.	100	10	5	2015.12
Pay (Platform)	Mandatory	Satellite (Direct operating channel)	2012.1.	2013.1.	70	7	4	2016.12
	Announcement Obligation	SO(Local channel)	2012.1.	2013.1.	70	7	4	2016.12
Pay (Program Provider)	Mandatory	News/ General service PP	2012.1.	2013.1.	100	10	5	2016.12
	Announcement Obligation	General PP IPTV CP	2012.1.	2013.1.	70	5	3	2016.12

Background

❖ Conventional DVS Production

- A professional DVS describer writes the script, then
- A broadcasting producer re-make the program using dubbing recorded by voice actors according to the scripts.

→ *It takes generally over 24 hours/program and costs to employ professional manpower practically.*

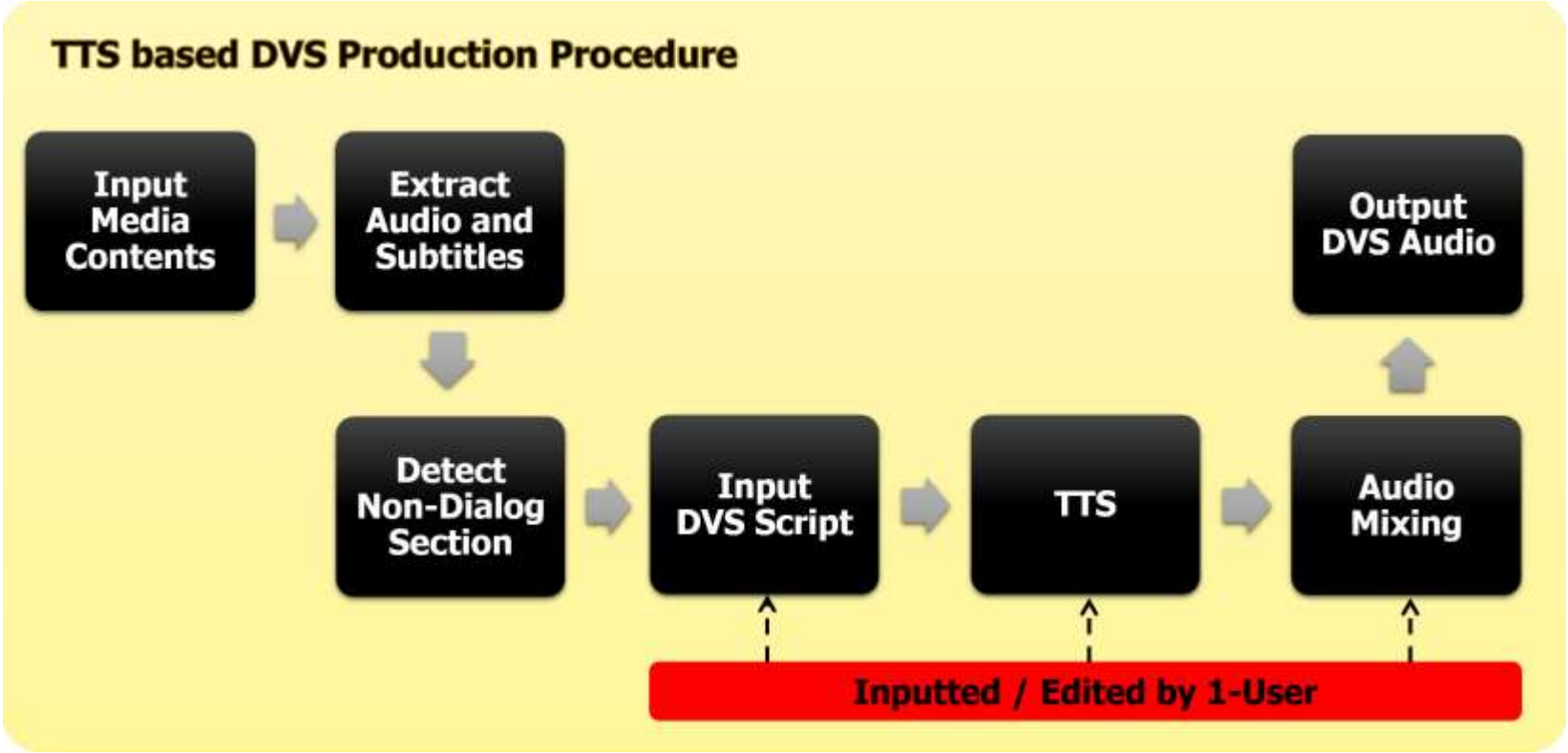


The Proposed - DVS using TTS

❖ To improve the practical limitation for DVS authoring

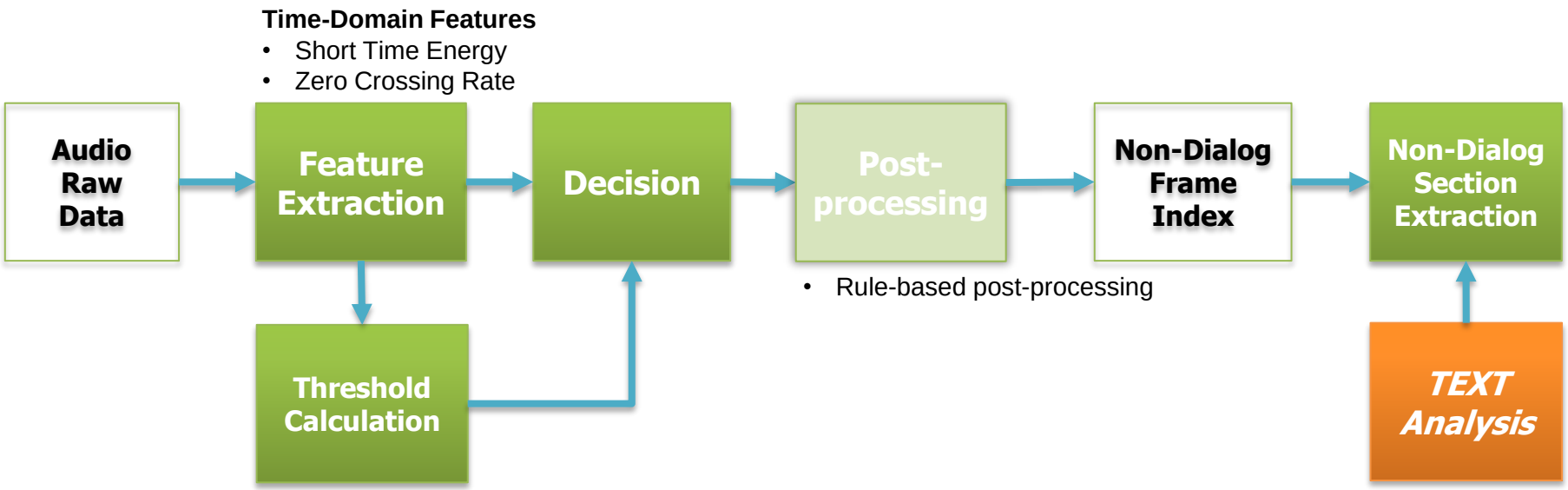
DVS Writer	✓ Recommendation for AD insertion (Non-dialogue section, length of description) ✓ Entering the DVS scripts
Voice Actor	✓ Text-To-Speech
Producer	✓ Mixing "Descriptive Audio" with "Mater Audio" ✓ Generating DVS contents

The Proposed - DVS using TTS



TTS: Text-To-Speech

Non-Dialog Section Detection from Audio



❖ Features

- Short-Time Energy (STE)
  - Normalized sum of squares of samples for each frame
- Zero Crossing Rate (ZCR)
  - Number of times that the time domain signal changes its sign

Non-Dialog Section Detection from Audio

❖ Decision

- Comparison between these values and their thresholds
  - Threshold: average value of each feature

❖ Post-Processing

- Rule based post-processing to fulfill the smoothing task
- Result of analyzing the conventional DVS contents
  - The length of audio description is over 1 s.
  - Exceptional case: the place name for the changed scene case.
- Considering the natural sound connection, the length of minimum valid non-dialog section is set as 2s.
  - At implementation, the non-dialog sections whose length are shorter than 2s were ignored and other non-dialog sections are recommended.

Non-Dialog Section Detection from Subtitles

- The terrestrial digital broadcasting services with MPEG-2 TS in Korea
  - Subtitle text data
    - Picture User Data inside Video Packet
  - Time information
    - PTS of related PES header

- TS: Transport Stream
- ES: Elementary Stream
- PTS: Presentation Time Stamp

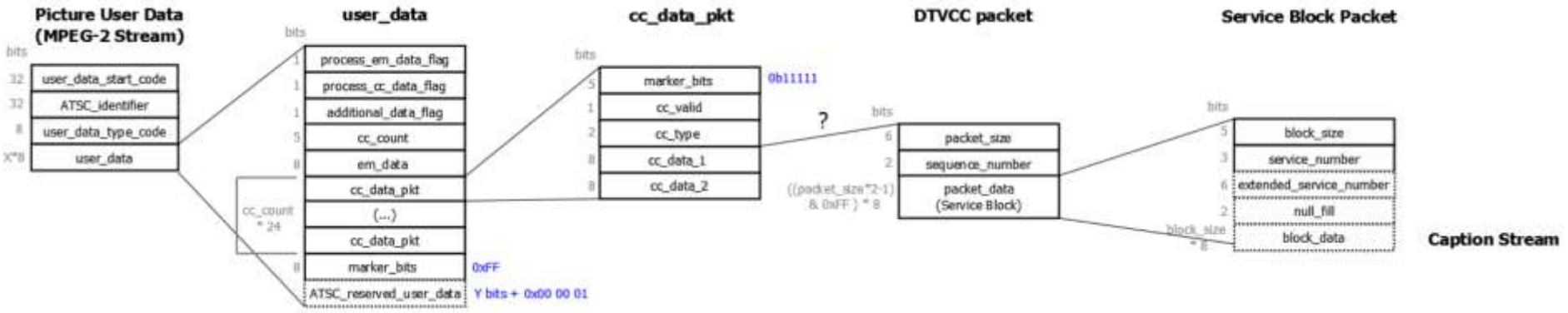
1. Recognition of one sentence

✓ To detect the punctuation marks which indicate the end of the sentence (.?!)

2. Extraction of its PTS

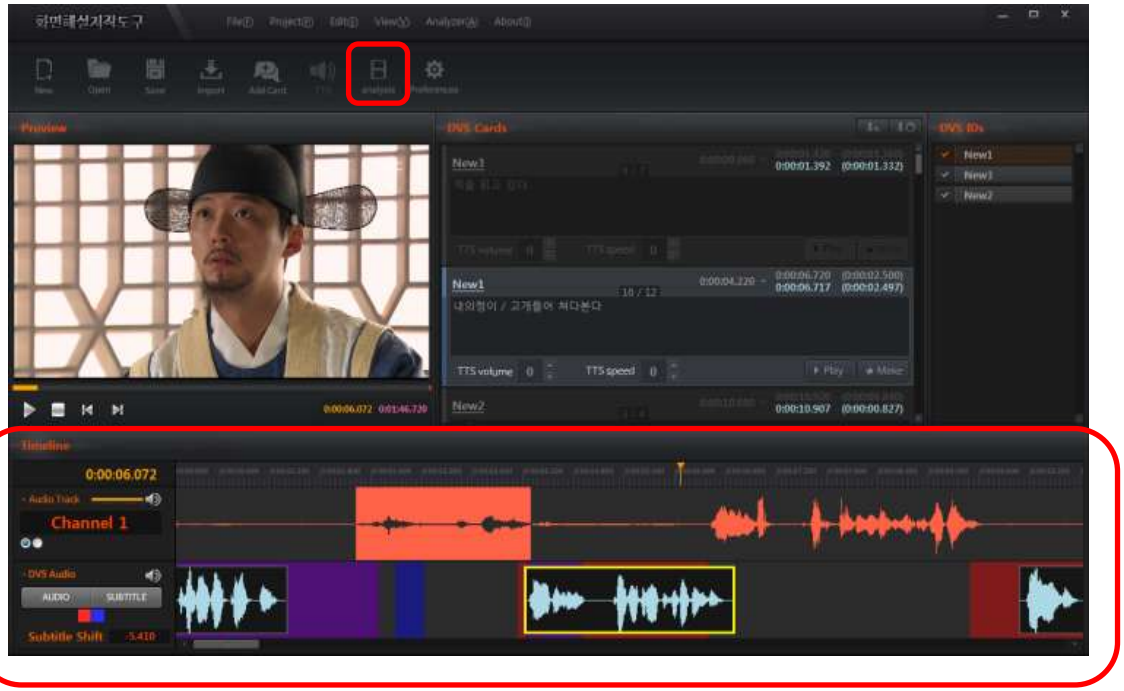
a. One is the PTS of TS packets including first character of the sentence

b. The other is the PTS of the TS packet including the punctuation

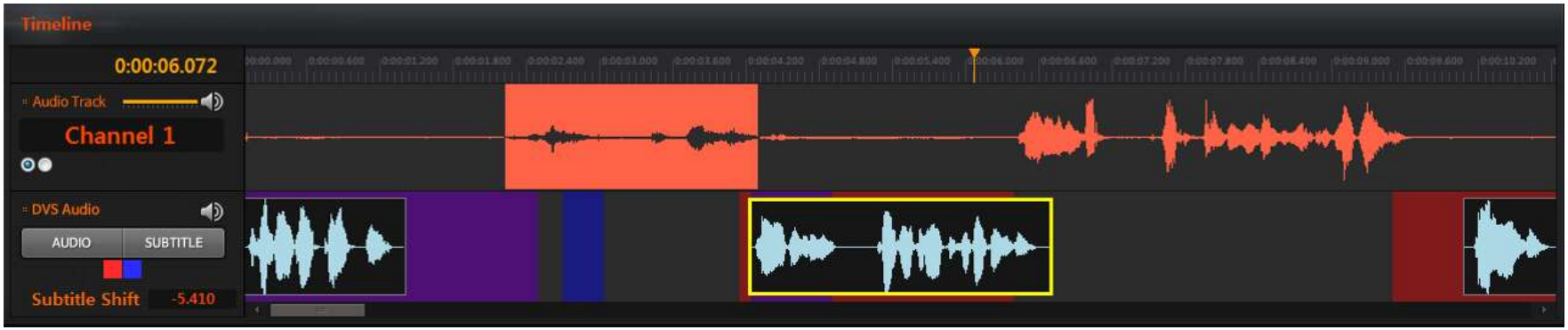


Example of the Implementation

- **The GUI (Graphic User Interface) of the proposed semi-automatic DVS authoring tool**
  - The imported media contents are captured from the real broadcasting with subtitles
  - Video and audio waveform are displayed on the ‘Preview’ and the ‘Audio Track’ window, respectively



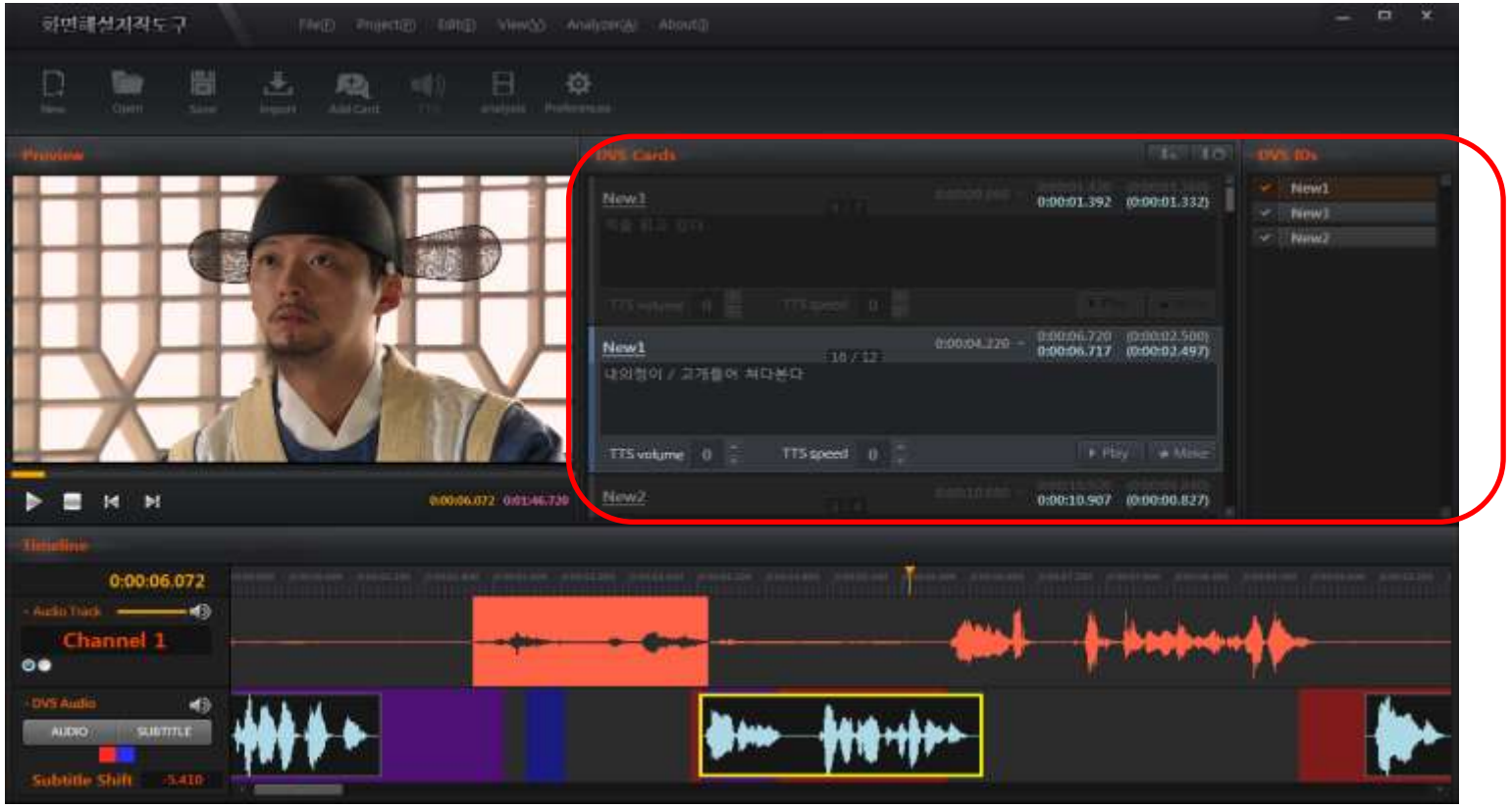
Example of the Implementation



- **DVS Audio**
  - TTS Audio waveforms are displayed
  - Red area: non-dialog section using audio analysis
  - Blue area: non-dialog section using subtitle analysis
  - Violet area: overlapped section of both audio and subtitle analysis
  - Yellow line (highlighted) box: TTS Audio waveform of focused ‘DVS Card’

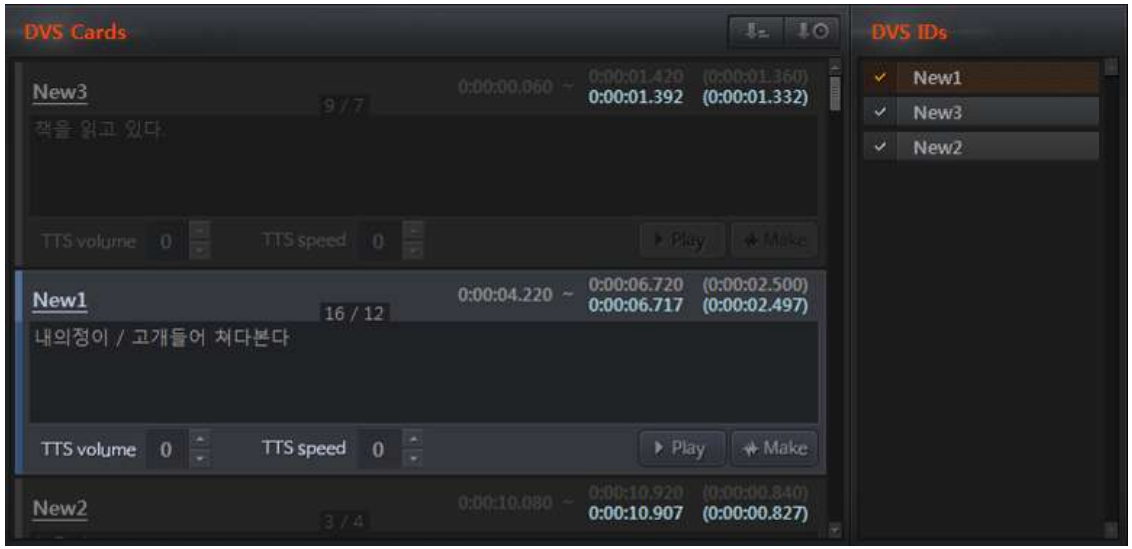
Example of the Implementation

- DVS scripts are able to be written in the ‘DVS Card’ located in the center of the authoring tool.



Example of the Implementation

- **DVS Cards**
  1. To enter the AD script
  2. To listen the TTS audio by clicking the ‘Play’ button
  3. To Control the TTS volume and speed
  4. To generate appropriate TTS AD by clicking the ‘Make’ button
     - TTS: Power TTS of Diotek
  5. Waveforms of generated TTS audio are displayed on the ‘DVS audio’ on the each ‘DVS Card’, respectively.



Conclusion

- **A semi-automatic DVS authoring method**
  - ✓ Non-dialog sections based on the audio/subtitles analysis are recommended
    - They are candidate sections where ADs can be entered.
  - ✓ Referring to the suggestions, appropriate AD scripts can be made
  - ✓ Synthesized ADs are generated using TTS
  - ✓ Mixing ADs with Master audio

- **Through the proposed, DVS contents can be generated semi-automatically and easily by one-author**

- **Currently, we have developed the trial version of the proposed and we proceed with adding more advanced functions and making its UI more convenient.**

Thank you.

Any Question?

jinsn@etri.re.kr