Mobile audio formats

Jarno Seppänen

Nokia Research Center

Contents

I. Overview

II. Natural audio formats
- 3GPP standard formats (AMR and AAC)
- Other formats (ADPCM, RealAudio, WMA, ...)
- 3GPP file format
- Standardization

III. Synthetic audio formats
- MIDI background
- SP-MIDI
- Mobile DLS and Mobile XMF
- Other formats (SMAF, i-Melody, ...)
- Content creation and tools

Natural audio and synthetic audio

- Two mobile audio format concepts: natural and synthetic audio
- Natural audio
  - E.g. MP3, WAV
- Synthetic audio
  - E.g. MIDI, text-to-speech (TTS)

	Band-width	CPU/ memory	Represen tation	Play control	Source material
Natural audio	20–200 kbps	Medium	Signal based	Limited (play/stop)	Acoustic recordings
Synthetic audio	1–10 kbps	Medium	Event based	Advanced	Musical performance

3GPP standard audio formats

- **AMR.** Mobile baseline speech
- **AMR-WB.** Wideband (7 kHz) extension to AMR
- **AMR-WB+.** Mobile speech and music
- **MPEG-2 AAC.** High-fidelity audio
- **MPEG-4 AAC+.** Higher coding efficiency for AAC
- **EAAC+.** Mobile speech and music
- **SP-MIDI.** Mobile synthetic audio with no sampled sounds
- **Mobile DLS.** Sampled sounds for wavetable synthesis
- **Mobile XMF.** Mobile synthetic audio

Other mobile audio formats

- **G.711 A-law, G.711 μ-law.** Speech coding by scalar quantization
- **G.726 ADPCM, IMA ADPCM.** Generic coding by adaptive quantization
- **GSM 06.10.** Legacy GSM speech codec
- **MPEG-1 Audio Layer III (MP3).** Predecessor to AAC high-fidelity audio
- **RealAudio (proprietary).** Generic speech and music
- **Windows Media Audio (proprietary).** Generic speech and music
- **MOD (proprietary).** Synthetic audio
- **SMAF (proprietary).** Synthetic audio
- **i-Melody (proprietary).** Synthetic audio

Natural audio formats

Mobile audio formats

AMR and AMR-WB Adaptive Multirate speech

- AMR (1998) and AMR-WB (1999) mobile speech codecs introduced in 3GPP standards
  - Adaptive multirate codec
- Very high sound quality for speech signals
- AMR speech bandwidth 3.5 kHz
- AMR-WB speech bandwidth 7 kHz
- AMR bit rates 5–12 kbps
- AMR-WB bit rates 7–24 kbps
- Mono channels only
- *.AMR and *.AWB files

AMR-WB+ speech and music

- AMR-WB+ introduced in 3GPP Release-6 standard (2004)
- AMR-WB+ is an audio extension to AMR-WB
- Very high sound quality for both speech and music signals
- Backwards compatible with AMR-WB
- High audio bandwidth (24 kHz)
- Mono and stereo channels
- Bit rates 14–48 kbps

AAC Advanced Audio Coding

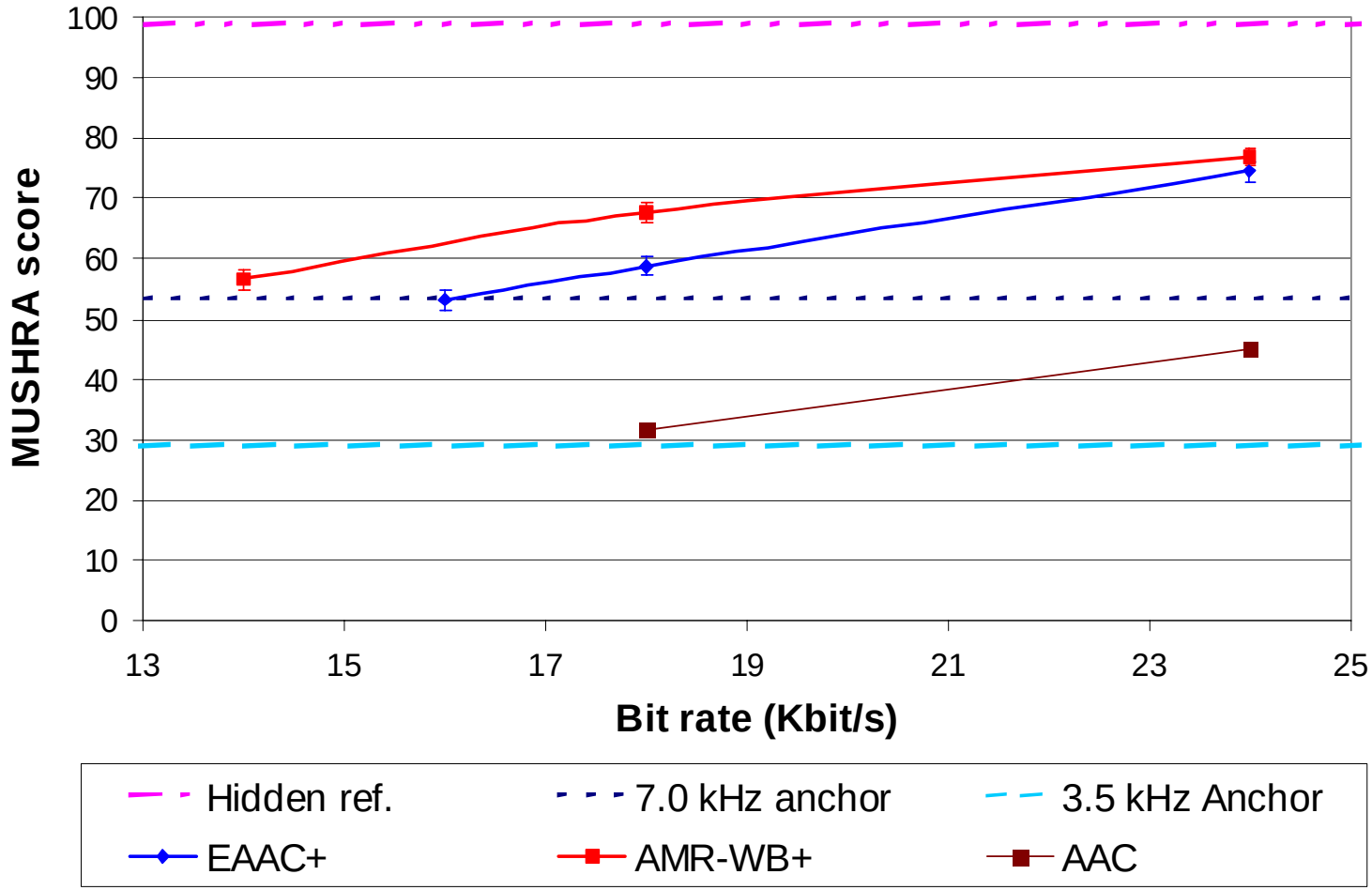
- Advanced Audio Coding (AAC) is part of MPEG-2 standard (1994)
  - Also known as MPEG-2 AAC
- Very high sound quality for music signals
- High efficiency AAC (HE-AAC) introduced in MPEG-4 standard (1998)
  - Based on MPEG-2 AAC with artificial bandwidth expansion
  - More efficient than MPEG-2 AAC
  - AAC Low Complexity profile (AAC-LC)
  - AAC Long-Term Prediction profile (AAC-LTP)
  - Also known as MPEG-4 AAC
  - *.M4A or *.MP4 files
- Enhanced AAC+ (EAAC+) introduced in 3GPP Release-6 standard (2004)
  - Based on HE-AAC with parametric stereo processing

Comparison of AMR-WB+ and EAAC+

- 3GPP have ordered extensive listening tests for AMR-WB+ and EAAC+ comparison
- Subjective listening tests
  - Stereo listening
  - Original PCM signal and two bandlimited signals used as anchors
- Content selected for typical mobile use scenarios
  - Eight music clips
  - Eight speech clips
  - Four clips of speech between music
  - Four clips of speech over music
- Test results on next slide

- See Mäkinen, J., *et al.*, *"AMR-WB+: A new audio coding standard for 3rd generation mobile audio services,"* submitted to ICASSP 2005

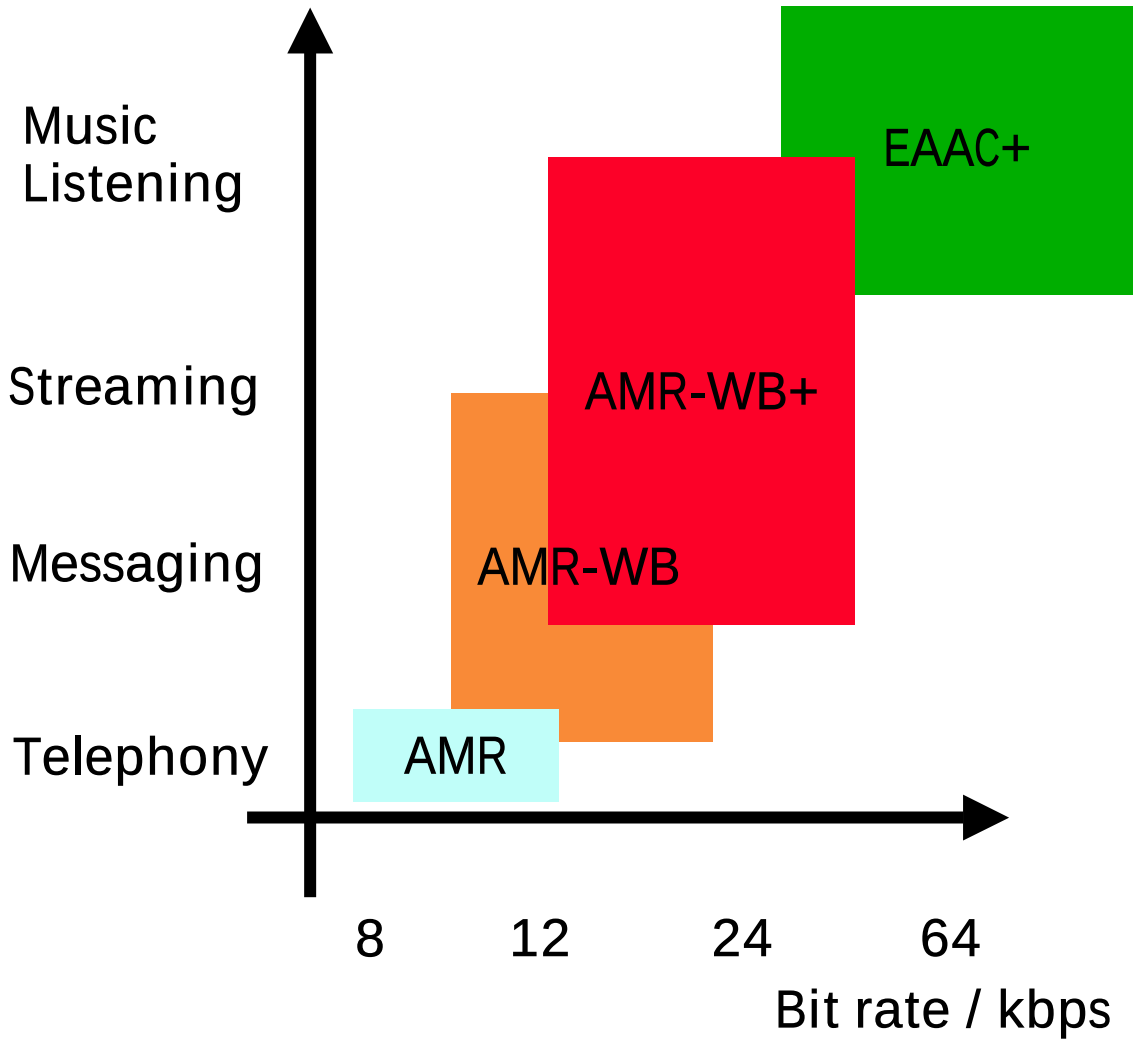
Comparison of AMR-WB+ and EAAC+



NOKIA

Natural audio applications

- Range of audio services
  - Telephony
  - Messaging
  - Streaming
  - Music listening
- Different codecs suit different purposes
- Also consider decoder CPU and memory consumption



Other standard formats

- G.711 A-law and μ-law
  - International Telecommunications Union ITU-T standard (1988)
  - Speech coding, logarithmic scalar quantization
  - Bit rate 64 kbps
- G.726 ADPCM, IMA ADPCM
  - International Telecommunications Union ITU-T standard (1990)
  - Bit rates 16–40 kbps
  - Generic signal coding by Adaptive Differential Pulse Code Modulation
- GSM 06.10
  - European Telecommunications Standards Institute (ETSI) standard (1992)
  - GSM standard speech codec
  - Bit rate 13 kbps
- MPEG-1 Audio Layer III (MP3)
  - Predecessor to AAC high-fidelity audio

Nonstandard formats

- RealAudio
  - Generic speech and music coding
  - First release in 1995
  - Proprietary low bitrate codecs
  - Uses AAC for high bitrates
  - Propriety of RealNetworks, Inc.
- Windows Media Audio
  - Generic speech and music coding
  - First release in 1999
  - Version 9 is comparable to AAC in quality
  - Propriety of Microsoft Corp.

3G mobile standardization: background

- UMTS standardization in 3GPP Third Generation Partnership Project
  - Creates 3G specifications and maintains GSM specifications
  - Not unlike a massive bureaucratic conglomerate
- U.S.-driven cousin organization for CDMA markets called 3GPP2
  - Competing with Europe-driven UMTS markets
- Multimedia codecs are one area of 3GPP standardization
  - 3GPP mandates stuff like radio interfaces, network protocols, data formats and codecs, and application services
  - E.g. Multimedia Messages (MMS)
- Latest approved codecs in 3GPP
  - AMR-WB+, EAAC+, Mobile DLS/XMF
  - H.264/AVC New MPEG video codec

3GPP file format

- 3GPP has its own file format (*.3GP)
  - Standardized in 3GPP Release-5 (2002)
- Packaging and distribution format for 3GPP standard content
- Lots of codecs: AMR variants, AAC variants, SP-MIDI, and Mobile DLS/XMF
- Mainly used for video content, not so much for audio
- Instance of ISO Base Media File Format
  - Also MPEG-4 format is derived from ISO Base Media File Format

3G mobile standardization: publications

- Public specifications available at http://www.3gpp.org
- AMR-NB
  - Specification at http://www.3gpp.org/ftp/Specs/html-info/26071.htm
  - Fixed-point source code at http://www.3gpp.org/ftp/Specs/html-info/26073.htm
  - Floating-point source code at http://www.3gpp.org/ftp/Specs/html-info/26104.htm
- AMR-WB
  - Specification at http://www.3gpp.org/ftp/Specs/html-info/26171.htm
  - Fixed-point source code at http://www.3gpp.org/ftp/Specs/html-info/26173.htm
  - Floating-point source code at http://www.3gpp.org/ftp/Specs/html-info/26204.htm
- AMR-WB+
  - Specification at http://www.3gpp.org/ftp/Specs/html-info/26290.htm
  - Fixed-point source code at http://www.3gpp.org/ftp/Specs/html-info/26273.htm
  - Floating-point source code at http://www.3gpp.org/ftp/Specs/html-info/26304.htm
- Enhanced AAC+
  - Specification at http://www.3gpp.org/ftp/Specs/html-info/26401.htm
  - Floating-point source code at http://www.3gpp.org/ftp/Specs/html-info/26410.htm
- Mobile XMF and Mobile DLS
  - Draft specifications at http://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_29/Docs/S4-030790.zip

Synthetic audio formats

Mobile audio formats

MIDI

- MIDI is a 20 years old synthetic audio standard
  - Musical Instrument Digital Interface
- **MIDI 1.0.** The base MIDI protocol
- **General MIDI (GM).** Common sound bank and controller definitions
- **DLS.** Downloadable sounds; sampled sound format
- **SMF.** Standard MIDI file; MIDI file format
- **XMF.** Extensible music format; container file format for MIDI
- Jointly developed by MMA and AMEI organizations
  - MMA – MIDI manufacturers' association (global)
  - AMEI – Association of musical electronics industry (Japan)

Standard MIDI files

- File format for storing MIDI protocol streams
  - In essence, musical performance data
- Does not specify sound synthesis method
- SMF cannot guarantee identical playback quality on different devices
  - Limited interoperability can be achieved with General MIDI sound bank
- Standardized in MMA and AMEI in 1991
- SMF consists of a time-stamped sequence of MIDI messages and meta-events
- Supports multiple parallel tracks
  - SMF format 0 means only one track
  - SMF format 1 means one or more tracks, played at the same time
- *.MID files

SP-MIDI background

- Scalable Polyphony MIDI (SP-MIDI) format for mobile applications
  - Previous General MIDI specifications are not scalable
  - Fixed CPU and memory requirements
- Standardized in
  - Music industry (MMA and AMEI)
  - Mobile industry (3GPP Release-5)
  - Java (J2ME) standardization
- The same content format for different complexity levels and devices
  - 3GPP profile defined for 5–24 voice polyphonies
- Adjustable implementation complexity
  - Low polyphony in low end mobile devices
  - High polyphony in high performance products
  - Scalable to low polyphony in complex use scenarios
- Improved interoperability
  - Embedded playback requirements
  - Content creator creates scalable content

NOKIA

SP-MIDI terminology

- **Scalable polyphony** is a property of a musical arrangement such that its polyphony requirements can be adjusted by muting MIDI channels during playback.
- **Note Generator** refers to hardware or software needed to play one note, e.g., to respond to one Note On–Note Off message pair.
- **Polyphony Level** is the maximum number of notes that an SP-MIDI compliant playback device can play simultaneously.
- **Channel Masking** is muting a number of lowest-priority MIDI channels during playback, and only playing the highest-priority channels. Channel masking is controlled by the MIP message.
- **Maximum Instantaneous Polyphony (MIP)** defines the total *number of notes required for playback of the respective MIDI channel together with all higher-priority MIDI channels.* The MIP table stores a cumulative polyphony requirement for the MIDI channels.

SP-MIDI technology

- Key objective: flexible and portable polyphony handling
  - Support content simultaneously optimized for several polyphonies
- Scalability is achieved by one additional message
  - Maximum instantaneous polyphony (MIP) message

  ⇒ Polyphony judgements that affect the arrangement are made by the composer or content producer

  ⇒ Playback device knows whether it can play the content and to which amount, and plays one of the *composed* arrangements
- No automatic arranging of music for different polyphonies
- The polyphony handling design is relatively simple but powerful
  - Content creator defines the requirements for sufficient playback
  - Channel priority order
  - Channel MIP values
  - Different playback devices can play different polyphony levels
- SP-MIDI defines a minimum sound bank
  - Subset of General MIDI sound bank
- SP-MIDI defines phone vibra control
  - Embed vibra on/off messages in ring tones

Scalability in SP-MIDI

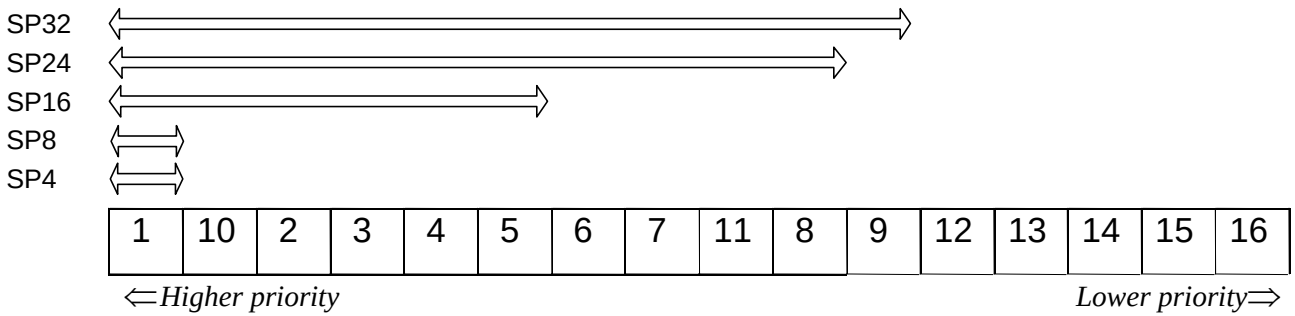
- MIP information is a table of 16 *{channel, polyphony}* value pairs
  - Stored in the beginning of an SP-MIDI song
- MIP information defines max. 16 different arrangements of the song
  - Each arrangement is defined by a group of channels
- Each arrangement has a required polyphony, as stored in the MIP table
- The playback device plays the arrangement for which it has sufficient polyphony
- Different players may and will play different arrangements of the same song
- Even a single player may play a different arrangement in different situations, depending on e.g. the availability of CPU or memory resources

Scalability in SP-MIDI

- Example SP-MIDI ring tone

Name	Channel	Number of Voices	MIP
Piano	1	4 Voices | 3 Voices	4
Drums	10	2 Voices | 5 Voices | 5 Voices	9
Bass	2	1 Voice | 1 Voice	10
Guitar	3	2 Voices | 3 Voices	12
Synth pad	4	3 Voices | 4 Voices	12
Orch hit	5	1 Voice	13
Organ	6	6 Voices	19
–	7	(none)	19
Percussion	11	4 Voices	23
Alien FX	8	3 Voices	23
Guitar 2	9	3 Voices	26
–	12	(none)	26
…	…	…	…

- Here, the MIP table is {4, 9, 10, 12, 12, 13, 19, 19, 23, 23, 26, 26, ..., 26}



DLS Downloadable Sounds

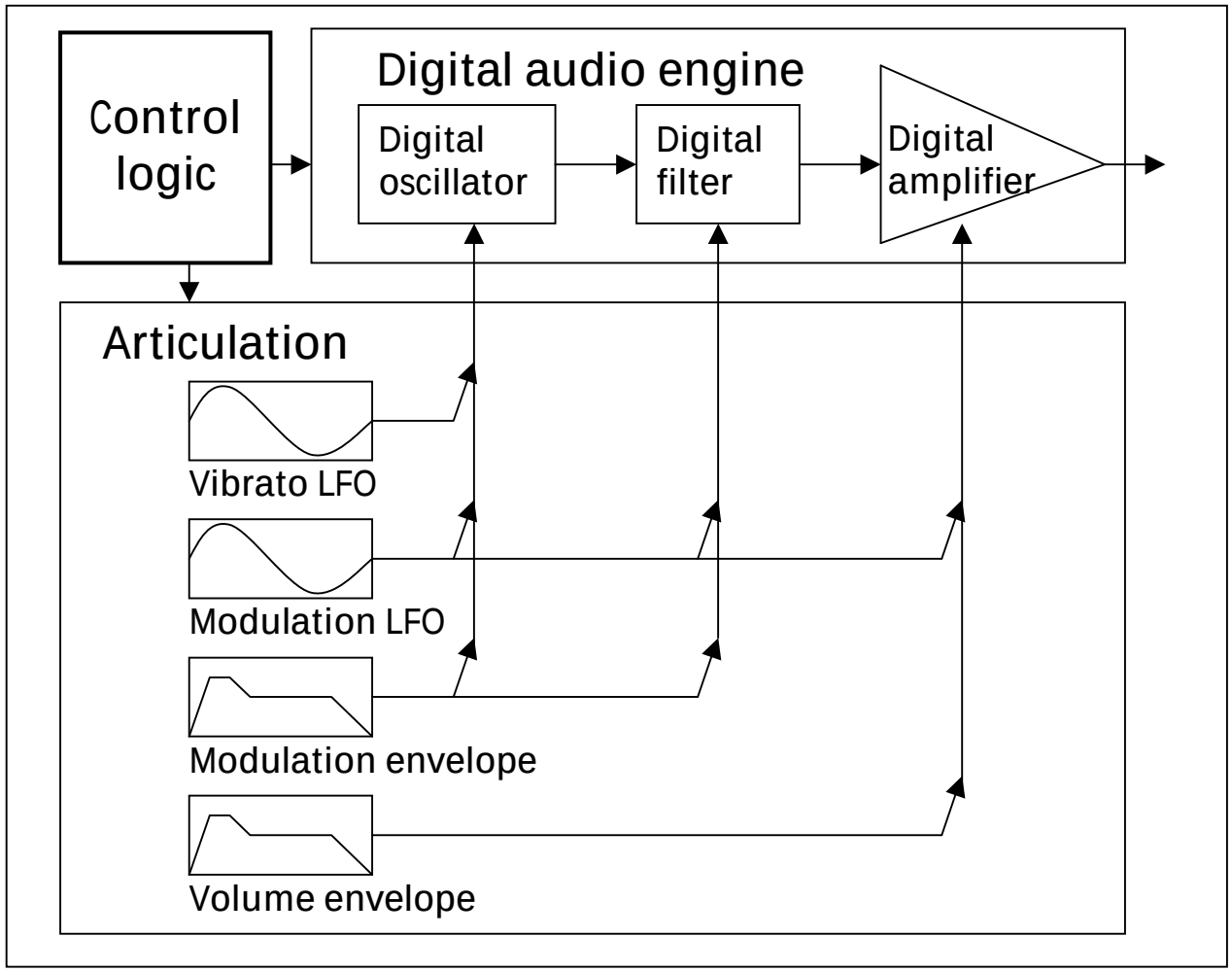
- DLS1 (1997), DLS2 (1999), and Mobile DLS (2004) standards from MMA
  - Originally based on proprietary SoundFont technology (Creative Labs)
- Designed for storage of sampled instrument sounds
- Based on wavetable synthesis paradigm
  - Does not support e.g. FM synthesis or virtual modeling synthesis
  - Contains articulation information in addition to the actual wavetable samples
  - Uses Microsoft RIFF file format
- Advantages
  - Common playback experience
  - Unlimited palette of possible instrument sounds (unlike General MIDI)
  - Compact and flexible music representation compared to sampled audio
- DLS2 adopted for MPEG-4 structured audio
  - Structured Audio Sample Bank Format (SASBF)
- DLS2 supported in Microsoft DirectX

Wavetable synthesis

- Wavetable synthesis is the most popular sound synthesis technology
- In addition to mobile devices, also widely used in sound cards, professional synthesizers, and instruments
- Recorded sounds or e.g. sine waves are stored to memory and read back during the synthesis
- Sounds can be modified and adjusted in real time during playback
- Computationally efficient but consumes memory for wavetable sound bank
- Every pitch of an instrument does not have to be stored in the memory since adjacent samples are similar to each other
- The missing pitches can be interpolated from the stored waveforms
- The stored wavetable is resampled for pitch shifting
  - Multirate filtering techniques used in audio sample rate converters are not used
  - Fractional delay filters support time-varying resampling in arbitrary ratios

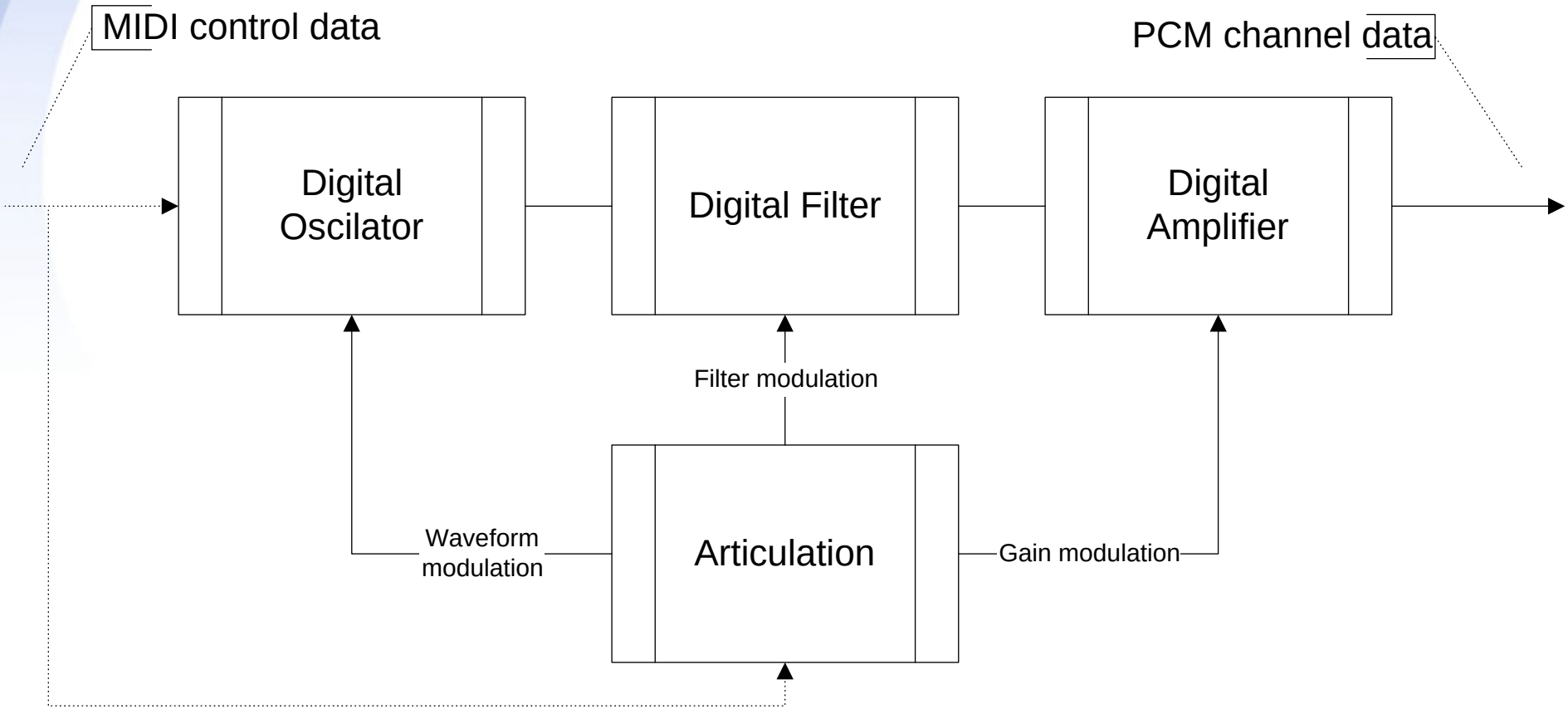
DLS sound synthesizer

- Wavetable synthesis architecture



DLS voice

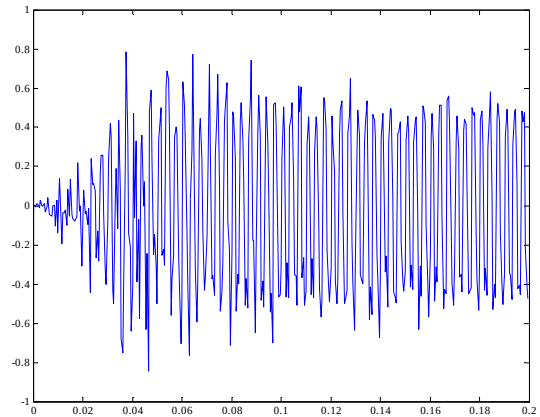
- The voice is the element of sound generation

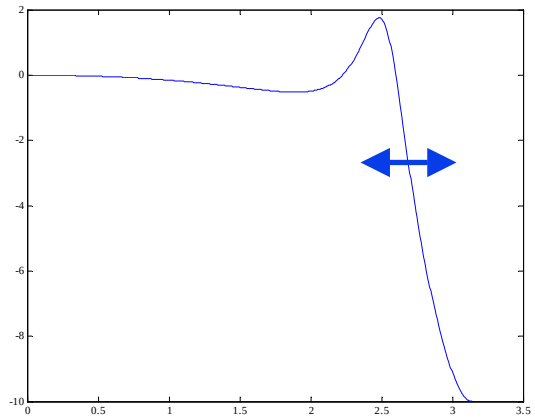


DLS voice

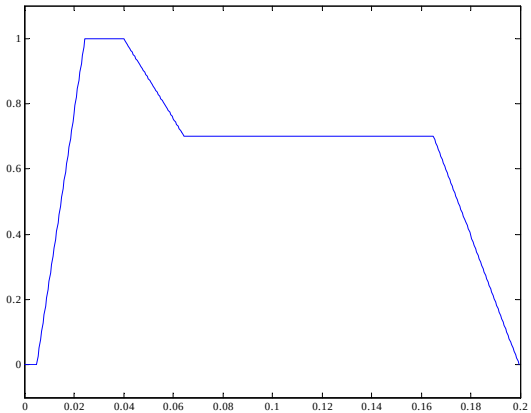
- Wavetable interpolation, filtering, and amplification

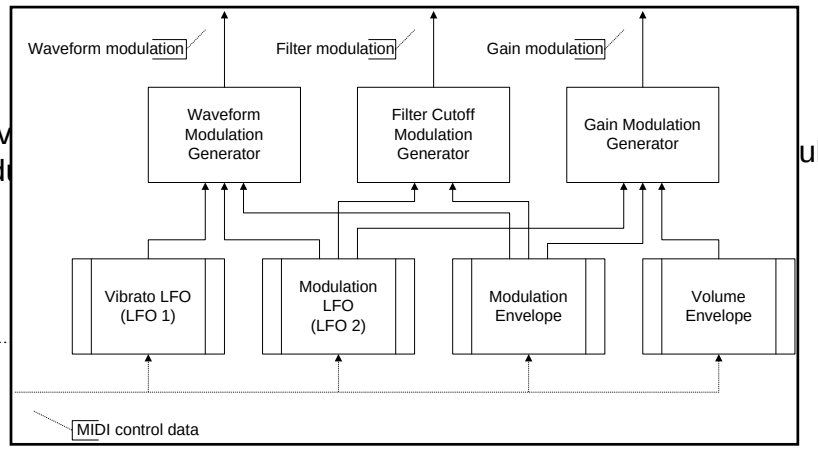
MIDI control data





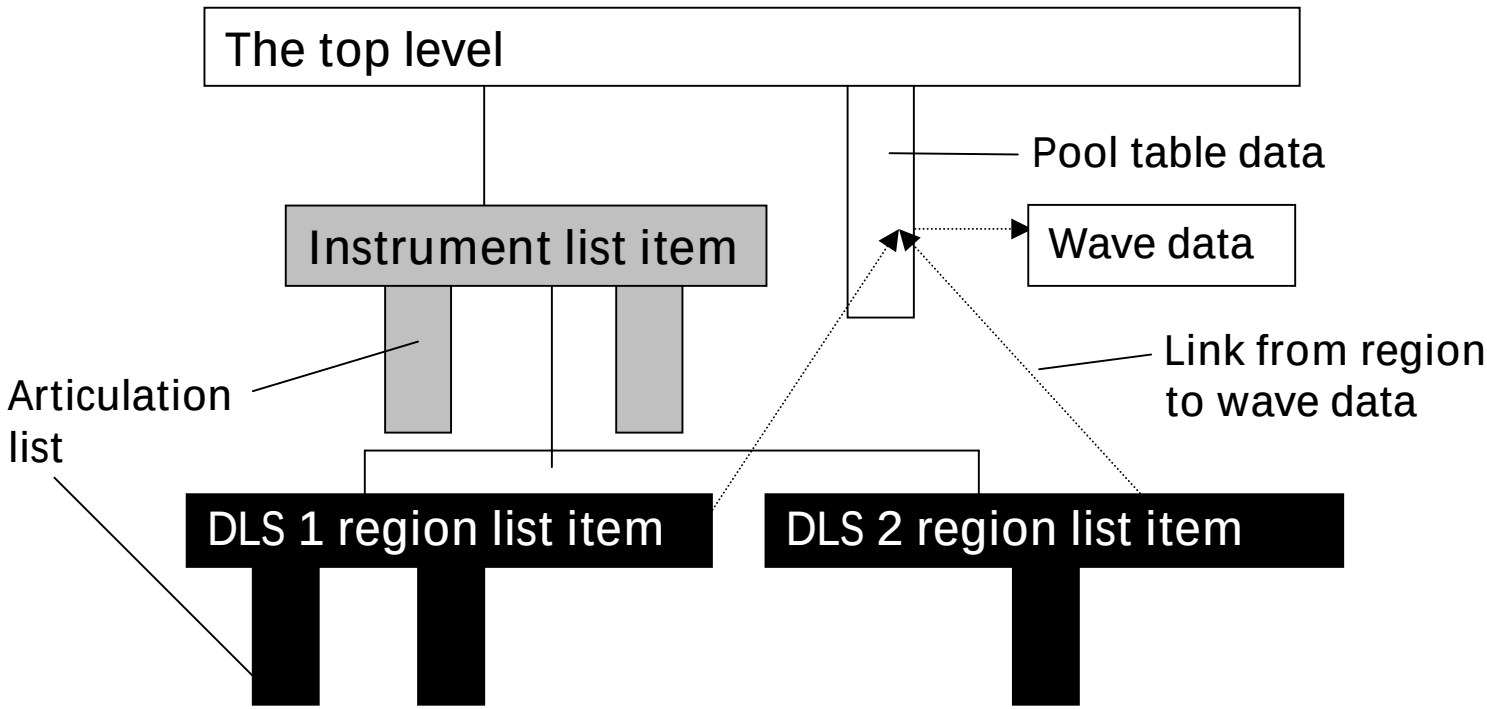
PCM channel data





DLS file format

- Basic data in four objects
  - Instruments, regions, articulation, and wavetable data
- Uses Microsoft RIFF file format



Mobile DLS

- Standardized in MMA/AMEI and 3GPP (2004)
  - Part of 3GPP Release-6
- Sampled wavetable synthesis
  - Technically based on DLS 2
  - Supports two different voice profiles
  - Relaxed implementation and memory requirements
  - Harmonization of DLS and SP-MIDI technology
- Unlimited sound bank, unlike General MIDI (GM)
- Support ITU-T G.711 A-law encoded wavetables
- Efficient real-time control
- *.MDLS files
- Vibra control with SP-MIDI
- Resource scalability with SP-MIDI

Mobile DLS profiles

Mandatory processing (MP)

- Wavetable sample sound source with loop points
- Two four-segment ADSR envelope generators
  - Attack–Decay–Sustain–Release
- One Low Frequency Oscillator (LFO)
  - Modulation LFO
- Total uncompressed Mobile DLS instrument size max. 7 kB

Optional processing (OP)

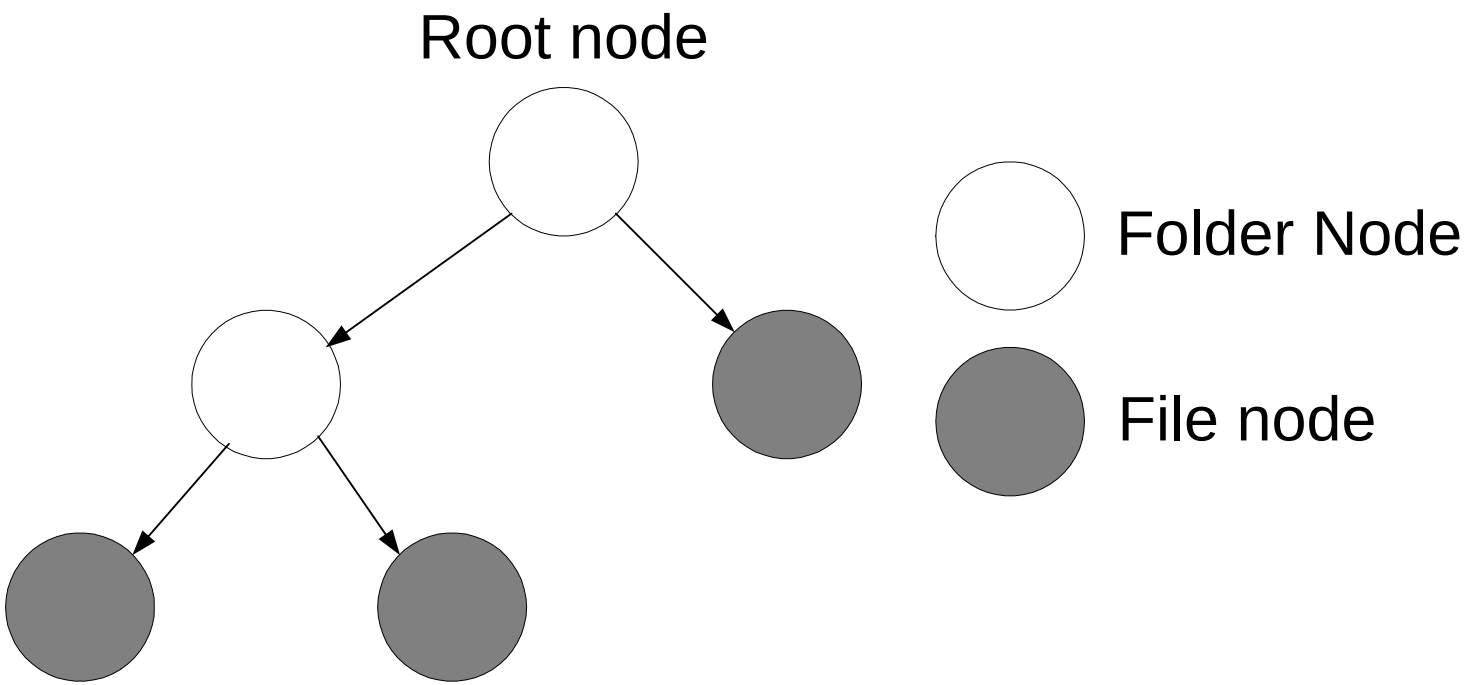
- Wavetable sample sound source with loop and release points
- Two six-segment DAHDSR envelope generators
  - Delay–Attack–Hold–Decay–Sustain–Release
- Two Low Frequency Oscillators (LFO)
  - Modulation LFO
  - Vibrato LFO
- Low pass filter with dynamic cutoff and resonance controls
- Identical to DLS2 synthesizer
- Total uncompressed Mobile DLS instrument size max. 15 kB

Mobile XMF

- Standardized in MMA/AMEI and 3GPP (2004)
  - Part of 3GPP Release-6
  - Based on proprietary Rich Music Format (RMF) technology (Beatnik)
- Container format for SP-MIDI and Mobile DLS content
  - Packaging synthetic audio content into a single file
- Designed for efficient distribution
  - Easier maintenance because the file format doesn't restrict the types of content
  - XMF allows compression and encryption methods for data members
- Compact representation for music
  - Extends from very small files to large files without limits
  - Variable Length Quantity (VLQ) number encoding
- *.MXMF files

Mobile XMF content

- Hierarchical file format in a tree structure
- XMF is a file-bundling format: an XMF file is a package of several files
- Contains all instrument, performance, copyright and other information of a song in a single file



SP-MIDI and Mobile XMF examples

	SP-MIDI	Mobile XMF	MP3
Bobby McFerrin *Don't worry be happy*	7 KB	92 KB	223 KB
	Polyphony 19	Polyphony 4	
The chemical brothers *Chemical beats*	23 KB	145 KB	518 KB
	Polyphony 10	Polyphony 3	
Darude *Sandstorm*	21 KB	169 KB	478 KB
	Polyphony 27	Polyphony 25	
Kraftwerk *The robots*	23 KB	228 KB	460 KB
	Polyphony 12	Polyphony 14	
Rage against the machine *Bulls on parade*	54 KB	113 KB	449 KB
	Polyphony 23	Polyphony 12	

NOKIA

Nonstandard formats

- MOD
  - Synthetic audio file format in PC Sound Tracker programs
  - Dozens of variations (MOD, XM, MED, S3M, ...)
- SMAF
  - Synthetic audio format
  - Propriety of Yamaha Corp. (Japan)
- i-Melody
  - Synthetic audio format
  - Propriety of NTT DoCoMo (Japan)

SMAF

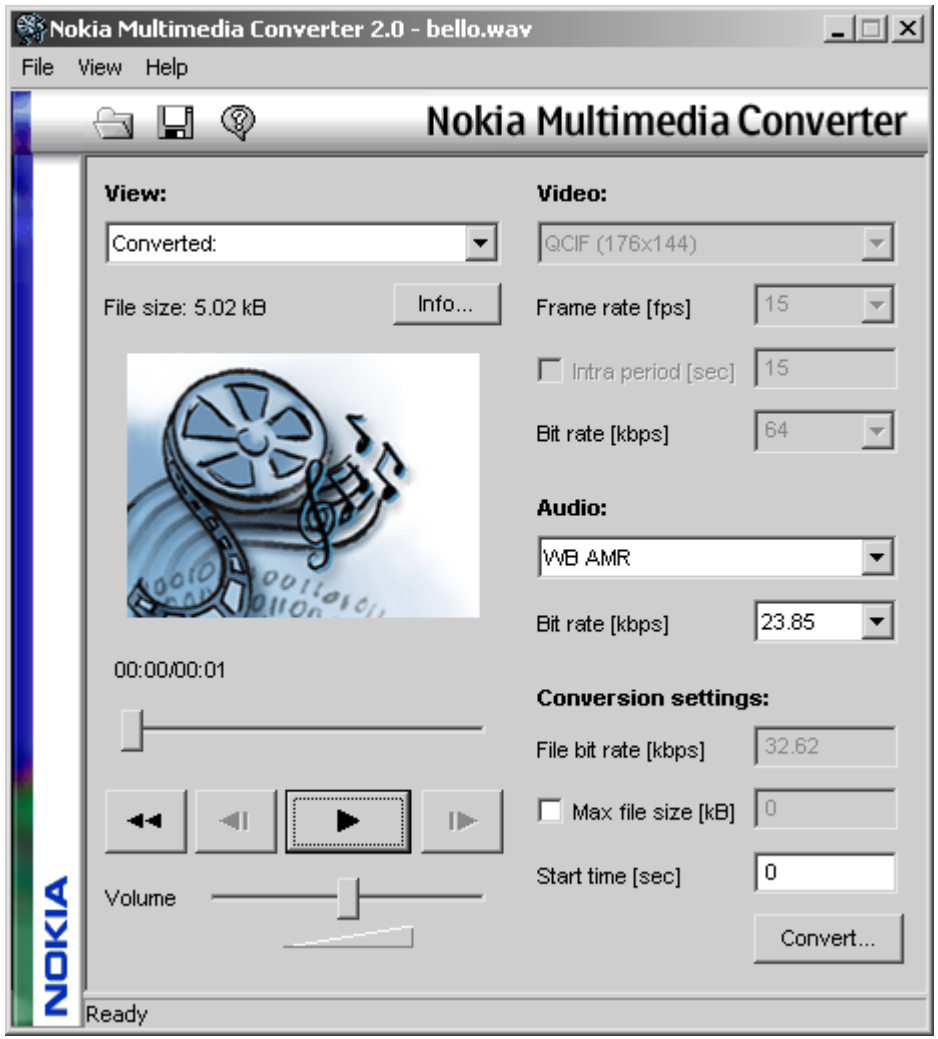
- Yamaha proprietary format for ring tones and mobile audio
- Mixed format
  - Synthetic audio (convertable from MIDI)
  - Natural audio (PCM and ADPCM)
  - Text
  - Graphics
- SMAF format linked to Yamaha SMAF player hardware products
- MA-1: four-polyphony frequency modulation (FM) synthetic audio
- MA-2
  - 16-polyphony FM synthesis
  - ADPCM natural audio
- MA-3
  - 32-polyphony FM synthesis
  - 8-polyphony ADPCM
- MA-5
  - 32-polyphony FM synthesis
  - 32-polyphony wavetable synthesis
- Geographically strongest in Asian markets (Korea, Japan)

Content creation

- Natural audio formats are encoded from ”raw” PCM/WAV audio signal
- Synthetic audio formats must be composed
  - General audio-to-MIDI conversion may not become feasible
- Nokia tools
  - Nokia Audio Suite
  - Nokia Multimedia converter
  - Nokia Lights Suite
  - Freely available from http://forum.nokia.com
- SP-MIDI content creation tools available from Forum Nokia
- Ensure interoperability by well-constructed multimedia content
- Available free of cost for all third party content creators
- Support VST professional music production technology

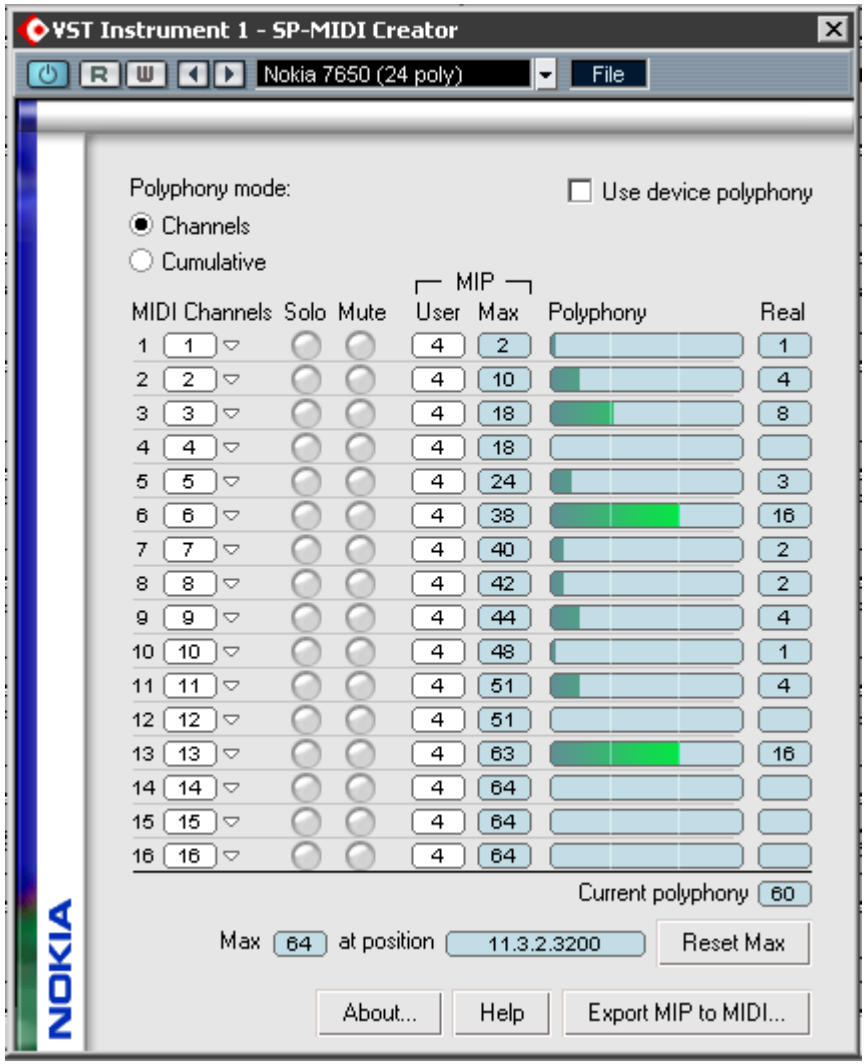
Nokia Multimedia Converter

- AMR, AMR-WB, and 3GP video content creation tool
- Convert PCM WAV audio files into AMR or AMR-WB format
- Changeable AMR bit rate
- Available from http://forum.nokia.com



Nokia Audio Suite

- SP-MIDI content creation tool
  - Real-time polyphony display
  - Real-time auralization
- Nokia Audio Suite can determine SP-MIDI MIP message during playback
- Polyphony and auralization settings changeable
  - According to phone model
- Plugs in to VST technology compliant audio applications
  - E.g. Steinberg Cubase
- Available from http://forum.nokia.com



Other tools

- Apple QuickTime: 3GPP, 3GPP2, MPEG-4 AAC, H.264, ...
- RealPlayer
- Helix player
  - Freely available
- Microsoft Windows Media Player
- Microsoft DirectMusic Producer
  - DLS creation
- Beatnik Mobile Sound Builder
  - Mobile XMF and SP-MIDI creation
  - Supports DLS formats
- Awave Studio
  - “Universal” audio file converter – supports over 240 file formats
  - Mobile DLS creation

Thank you!