ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Multimedia Technology

Section # 22: Synchronization

Instructor: George Xylomenos

Department: Informatics

Contents

- Introduction
- Synchronization requirements
- Resynchronization
- Distributed synchronization
- Synchronization specifications

Introduction

Class: Multimedia Technology, **Section # 22:** Synchronization

Instructor: George Xylomenos, **Department:** Informatics

What is synchronization? (1 of 3)

- Multimedia means unifying different media
  - Digital media representation
  - Independent media handling
  - So, we need to synchronize them!
- At which level do we synchronize them?
  - LDU: Logical Data Unit (of some medium)
  - Synchronization at the LDU level
    - Example: a video frame

What is synchronization? (2 of 3)

- Single medium synchronization
  - Applies to continuous media only
  - Timing relations between LDUs of the medium
    - Periodic playback, e.g., at 30 fps
- Multiple media synchronization
  - Timing relations between many media
  - One of them must be continuous
    - Example: subtitles, lip sync

What is synchronization? (3 of 3)

- Live synchronization
  - All media are digitized together
    - Example: audio and video in a camera
  - Need to reproduce their original relations
- Synthetic synchronization
  - Media are digitized or created independently
  - The relations are defined afterwards
    - Example: audio and subtitles in a movie

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Synchronization requirements

Class: Multimedia Technology, **Section # 22:** Synchronization

Instructor: George Xylomenos, **Department:** Informatics

Requirements (1 of 2)

- Single medium synchronization
  - Limit jitter during playback
    - Strictly periodic playback
    - Example: one frame every 40 ms (25 fps)
- Multiple media synchronization
  - Synchronization after independent delays
    - Each medium may face different conditions
    - Example: reading independent video and audio files

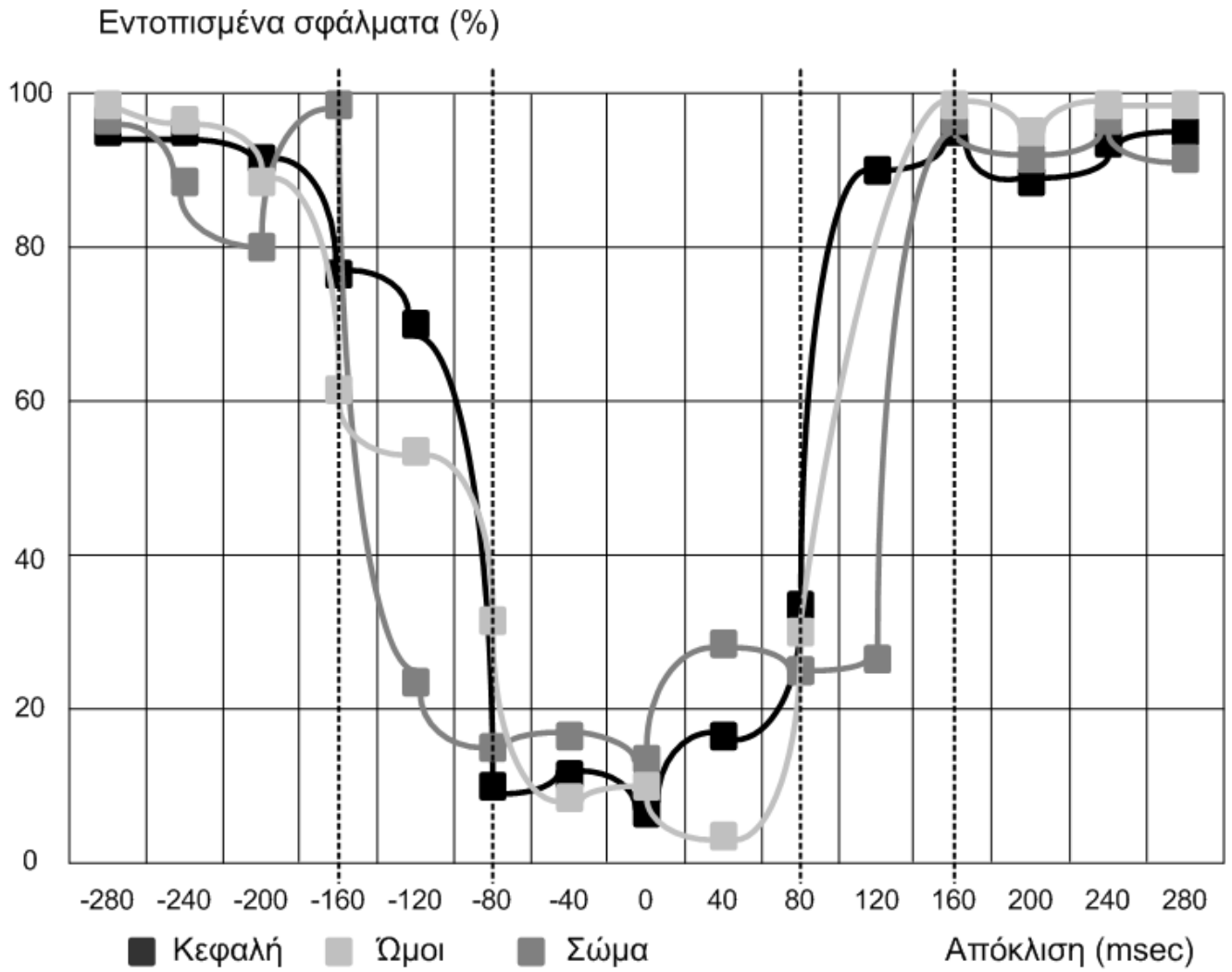
Requirements (2 of 2)

- How much tolerance do we have?
- Depends on multiple factors
  - Nature of the medium
    - Audio is more restrictive than video
  - Way media are combined
    - Lip sync is more restrictive than subtitle sync
  - User expectations
    - Sound usually follows video (why?)

Lip sync (1 of 4)

- Lip synchronization (lip sync)
  - Video (lips) and audio (voice)
  - Positive offset: video follows audio
  - Negative offset: audio follows video
- Empirical study based on questionnaires
  - Subject viewed a news presenter
    - Head, shoulders and body view
  - AV offset set at +/- k x 40 ms

Lip sync (2 of 4)



Lip sync (3 of 4)

- Tolerance to offset
  - Similar results (curves)
    - Head view makes offset more noticeable
  - Positive offset is more annoying
    - Curves rise faster on the right side
  - Negative offset is a natural phenomenon
    - Light travels faster than sound
    - For a distance of 34.3 m, negative offset ~100 ms

Lip sync (4 of 4)

- Three “areas” of synchronization
- In sync: –80 ms to +80 ms
  - Imperceptible or acceptable
- Out of sync: more than +/–230 ms
  - Perceptible and annoying
- Intermediate: +/- 80 to 230 ms
  - Acceptance depending on viewpoint

Pointer sync

- Example: a person comments on an image
  - We present an image
  - We point at various points
  - And comment on what we show
  - Up to 750 ms when the pointer follows the voice
  - Up to 500 ms when the voice follows the pointer
- Similar to subtitle sync

Audio channel sync

- Stereo audio: 11 µs (not ms!)
  - Mimics the small differences between our ears
  - Which help locate audio sources
- Speech with music on the background: 500 ms
  - Volume change in movies
- Discussion from different channels: 120 ms
  - Same as in natural conversations

Requirements table

Medium		Application	Tolerance
Video	Animation	Correlation	+/- 120 ms
	Audio	Lip sync	+/- 80 ms
	Image	Overlaid	+/- 240 ms
		Nov overlaid	+/- 500 ms
	Text	Overlaid	+/- 240 ms
		Not overlaid	+/- 500 ms
Audio	Animation	Correlation	+/- 80 ms
	Audio	Closely coupled (stereo)	+/- 11 µs
		Coupled (dialog)	+/- 120 ms
		Loosely coupled (background)	+/- 500 ms
	Image	Closely coupled (score)	+/- 5 ms
		Loosely coupled (presentation)	+/- 500 ms
	Text	Text commentary	+/- 240 ms
	Pointer	Pointer commentary	-500 ms / +750 ms

Synchronization levels

- Production level sync
  - Needed during processing
  - Tiny or no offset
    - Allows some relaxation during presentation
- Presentation level sync
  - Needed during playback
  - Offset within requirements
    - See previous table

QoS and QoE (1 of 2)

- Quality of Service (QoS)
  - Objective criteria
    - Example: latency limits
  - Relies on measurement equipment
- Quality of Experience (QoE)
  - Subjective criteria
    - Example: Likert scale for sync
  - Relies on questionnaires

QoS and QoE (2 of 2)

- In general, complex relationship
  - Media offset is part of QoS
  - Offset tolerance is part of QoE
  - Tolerance depends on many factors
- QoE functions
  - Combination of multiple QoS criteria
  - Example: streaming video QoE
    - QoE=f(number & duration of stalls, fps, resolution)

Resynchronization

Class: Multimedia Technology, **Section # 22:** Synchronization

Instructor: George Xylomenos, **Department:** Informatics

Why resync? (1 of 2)

- Variable access delays
  - Video and audio in different files
  - Seek times are unpredictable
- Variable network delays
  - Loss of sync for a single medium
    - Due to delay jitter
  - Loss of sync between media
    - Due to different delays

Why resync? (2 of 2)

- Gaps in the LDU sequence
  - Entire LDUs may be lost
    - Lost or delayed packets
    - If a packet misses its deadline, it is lost
- Prevention methods
  - Prefetching and buffering
  - Data interleaving
  - Forward error correction

Resync options (1 of 2)

- What if nothing works?
- Restricted blocking
  - Leave gaps for minor losses
    - Audio blanks out for a while
  - Repeat LDUs for larger losses
    - Freeze videos
  - This leads to perceptible issues
    - Audio clicks, video blocking

Resync options (2 of 2)

- Media resampling
  - Works for long-term issues
  - Slow down (and speed up) playback
  - Used to resync media when clocks drift
  - Repetition/skip/interpolation of LDUs
    - Also used in PAL/SECAM/NTSC/film conversions
    - Creation and dropping of frames
  - Needs considerable processing

Distributed synchronization

Class: Multimedia Technology, **Section # 22:** Synchronization

Instructor: George Xylomenos, **Department:** Informatics

Distributed multimedia

- Two main issues arise with networks
- Delay (or latency)
  - Time between capture and playback
  - Sampling, queueing, transmission, …
- Jitter (delay variance)
  - Mainly due to queuing
  - Main issue with packet switching

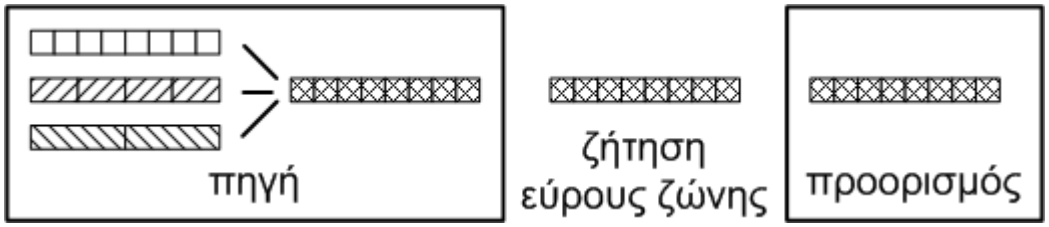
Transmitting specifications (1 of 2)

- When are specifications sent?
  - For media synchronization
- Before the presentation
  - Makes sense for synthetic sync
  - Adds some extra delay before media transmission
- Over a separate channel
  - Makes sense for live sync
  - Needs extra sync channel

Transmitting specifications (2 of 2)

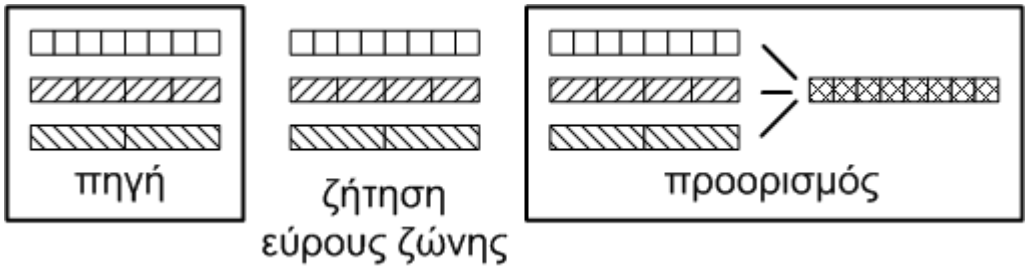
- Multiplexed with media
  - No initial delay
  - No extra channels
  - But, no independent media handling
    - Independence is lost during transmission

Synchronization location (1 of 2)



- At the source
  - All media synced at the source
  - One channel can be used for everything
    - Multiplexing of all media streams
  - All media face the same issues
    - Delay, jitter, loss

Synchronization location (2 of 2)

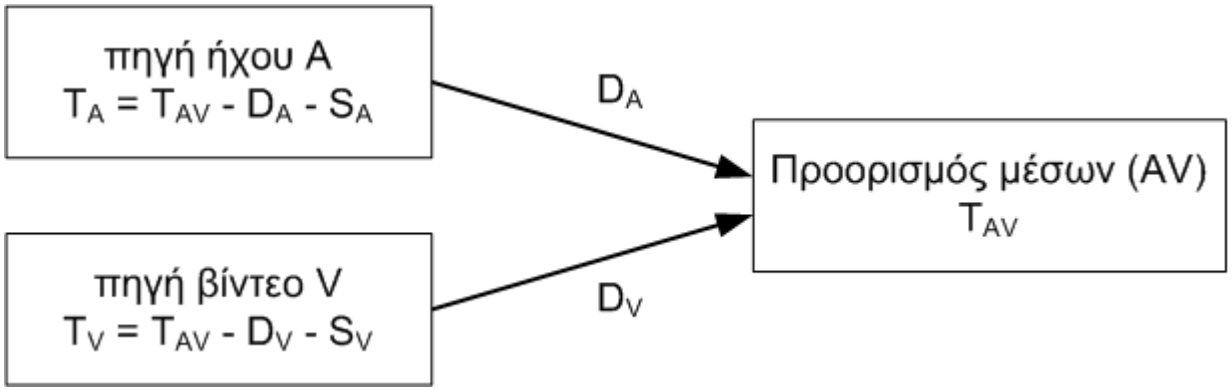


- At the sink
  - Media synced before playback
  - Independent media transmission
    - Each media may face different issues
    - Requires some buffering at the receiver
  - Allows changes at the sink

Synchronization basis

- How can we actually know what to do when?
  - Media can use a common time basis
  - Example: SMPTE timecode
    - Hour:minute:second:frame
    - Coded with 32 bits (using BCD)
  - But each device has its own clock
    - Must translate it to SMPTE
    - Still, clocks diverge from each other

Synchronization tolerance



- Say that audio and video are shown at T_{AV}
  - Audio: must start transmission at at $T_A=T_{AV}-D_A-S_A$
  - Video: must start transmission at $T_V=T_{AV}-D_V-S_V$
    - D_A, D_V: transmission delays
    - S_A, S_V: timing offset between clocks (unknown)
  - Need to estimate an upper clock offset
  - Also need to reserve buffers and plan in advance

Synchronization specifications

Class: Multimedia Technology, **Section # 22:** Synchronization

Instructor: George Xylomenos, **Department:** Informatics

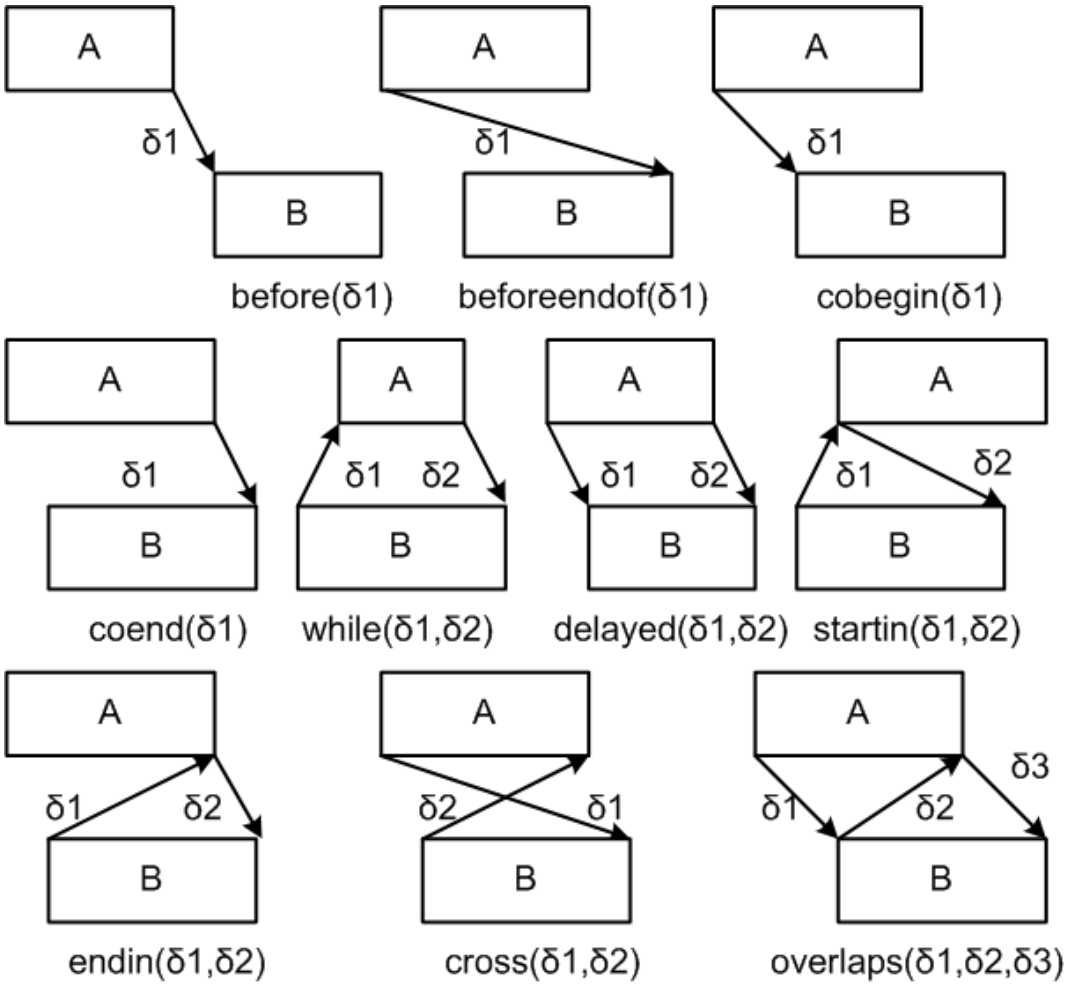
How to specify sync? (1 of 2)

- Live sync is straightforward
  - Must replicate timing during capture
  - Good for video and audio from same camera
- What about synthetic sync?
  - Possibly due to media independence
    - Video and subtitles
  - Or due to presentation complexity
    - Multipart interactive presentations

How to specify sync? (2 of 2)

- Sync specifications
  - We need a way to express specifications
  - How are media streams related?
  - Whan should an event happen?
  - How does it relate to other events?
  - How to account for interactivity?
  - We need a specification “language”

Interval-based (1 of 3)



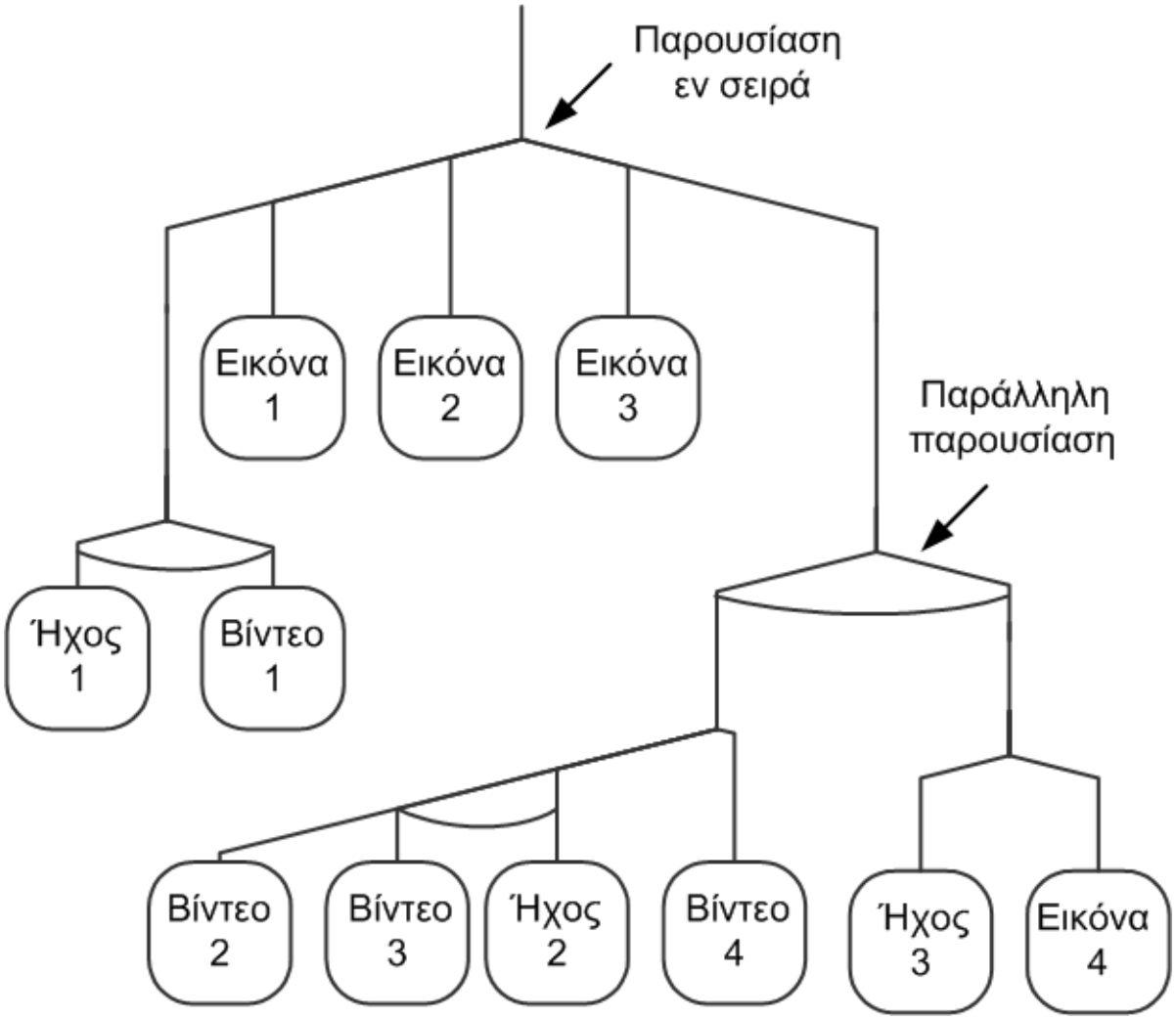
Interval-based (2 of 3)

- Interval-based specifications
  - Media streams are broken into intervals
    - Example: sound effect, video clip
  - Many types of interval relations
  - Showing how intervals A and B are related
  - Parameters δ1, δ2, δ3
  - Example: before(δ1)
    - Interval A begins δ1 before the beginning of B

Interval-based (3 of 3)

- Interval-based specifications
  - Allows unknown parameters
    - Can be filled in during the presentation
    - This allows dynamic navigation
  - Can only deal with entire intervals
    - May need to break down clips for example
  - Possible issues with unknown parameters
    - It may be impossible to achieve sync

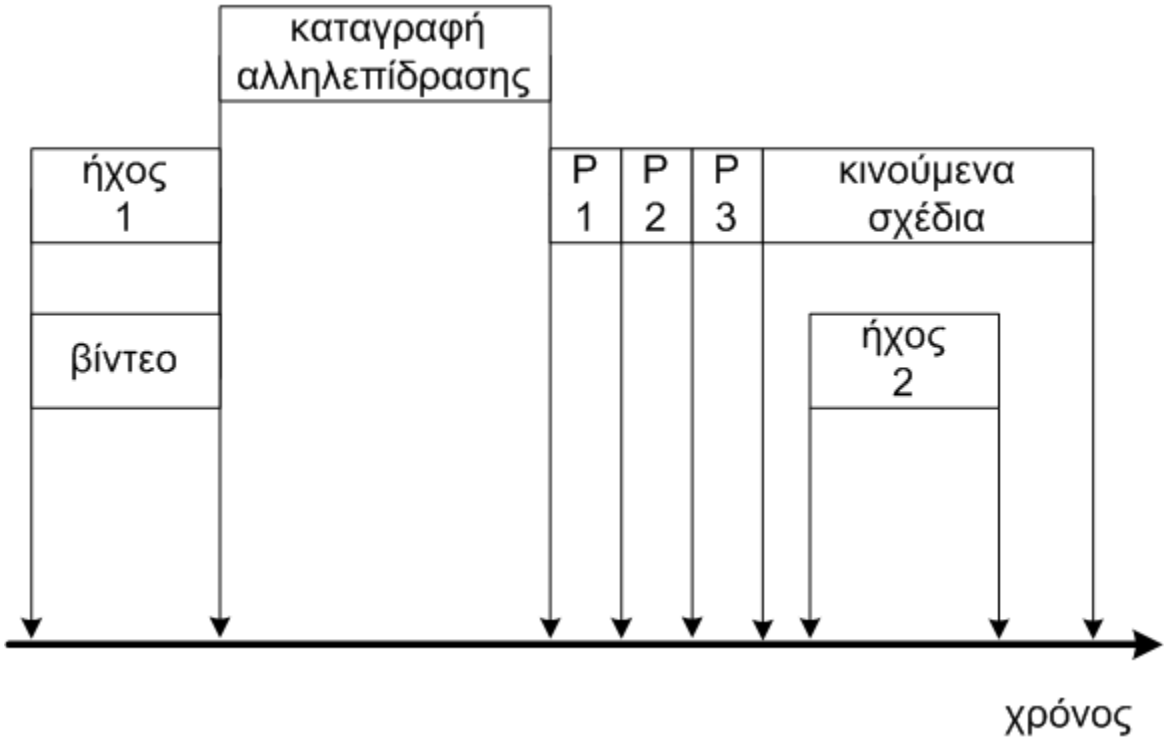
Hierarchy-based (1 of 2)



Hierarchy-based (2 of 2)

- Control flow-based specifications
  - Synchronization at specific control points
- Hierarchy-based specifications
  - Presentation in series or in parallel
    - Leaf: media object
    - Nodes: order of presentation
  - Unknown delays can be objects
  - Synchronization only at object beginning/end
  - Cannot express more complex relations

Axis-based (1 of 2)



- Axis-based specifications
  - Central axis used as a reference
  - Media are synced to this axis

Axis-based (2 of 2)

- Axis-based specifications
  - Cannot handle unknown durations
    - For example, waiting for a user to click
  - All media must be synced with the same axis
  - Usually, the audio clock is the axis
    - Audio is most sensitive to jitter
    - And its sampling clock is in the kHz range
  - Most common in applications

End of Section # 22

Class: Multimedia Technology, **Section # 22:** Synchronization

Instructor: George Xylomenos, **Department:** Informatics