COMP7503 Multimedia Technologies IntroductionDr. BillLuo COMP7503 2024-2025What isMultimedia? ❖Everybody thinks they know about the word “multimedia”, but most of them think of multimedia from narrow perspectives MultimediaIt’s for Entertainm ent It refers to musicIt refers to photo It refers to WebpageIt refers to videoIt refers to ebook COMP7503 2024-2025 2Is audio alone enough? Video adopted from ://ww w.youtube.com/watch?v=FeTOf49mUyA COMP7503 2024-2025 3Istextinformation alone enough? ❖Under which scenario a person would say the following ❖“Just lie down, shut up, and open wide” COMP7503 2024-2025 4 Why Multimedia? ❖How do humans receive information? ❖Five senses ❖Sight ❖Hearing ❖Taste ❖Smell ❖Touch ❖ ❖Each sense provides only partial information Multimedia refers to the combined use of several media that stimulate one or more of these five senses to deliver relevant information to users COMP7503 2024-2025 6Why Multimedia? (Cont’d) ❖Audio and Visual Information dominates human activities ❖Humans are equipped with powerful visual and audio analysis capability ❖Taste, smell, and touch cannot be easily acquired and reproduced by existing multimedia computing technologies COMP7503 2024-2025 7Then, What is computing ? ❖Computing is any goal -oriented activity requiring, benefiting from or creating computing machinery. ( ) ❖Multimedia computing means utilise signals from multifarious sensors and present to users only the relevant information in the appropriate sensory modality (media) ❖The goal of multimedia computing ❖Information sharing on experiences ❖Knowledge extracting from experiences COMP7503 2024-2025 8Communications in Human Society ❖ ❖ ❖ ❖The Communication is key for usto ❖Share experiences ❖Create, maintain, sustain, and propagate knowledge Communication across space ❖Allows people to exchange information independent of their current locations ❖Examples? What are the technologies required to support this? Communication across time ❖Allows people to experience an event over and over again without having to be there at the exact moment ❖Examples? What are the technologies required to support this? Communication to as many people as possible (addressee ) ❖Examples? What are the technologies required to support this? COMP7503 2024-2025 9Human Communication Means (1) ❖Spoken Languages ❖Consisted of analog sounds uttered with the speech -producing infrastructure in the throat ❖❖Able to convey complex facts Any invariance across time, space, and addressee? COMP7503 2024-2025 10Human Communication Means (2) ❖Written Languages ❖Evolved from drawings and paintings ❖Enables sharing of experiences and knowledge ❖Initial bulky preservation techniques (e.g. stone tablets) evolved into more practical storage devices and writing methods ❖e.g. Paper and ink ❖Any invariance across time, space, and addressee? Image adopted from COMP7503 2024-2025Image adopted from maher.blogspot.hk 11Human Communication Means (3) ❖Printing Press ❖Invented by Johannes Gutenberg around year 1440 ❖Enabled mass communication by making creation, storage, and distribution of documents more manageable ❖Makes information and knowledge sharing earlier ❖Any invariance across time, space, and addressee? Image acquired from -inside -the-gutenberg -printing -press/ COMP7503 2024-2025Human Communication Means (4) ❖Telegraph ❖Allows instantaneous communication of symbolic information over long distances ❖Any invariance across time, space, and addressee? ❖Telephones ❖Allows natural communication (talking) over long distance ❖Any invariance across time, space, and addressee? Image acquired from COMP7503 2024-2025Human Communication Means (5) ❖ ❖Radio ❖Allows distribution of sound wirelessly ❖First form of instantaneous mass communication ❖Any invariance across time, space, and addressee? Movies and Television ❖Combine our senses of sight and hearing for more natural communication ❖❖Effective way to key into our emotions Any invariance across time, space, and addressee?Image acquired from -us/ Image acquired from COMP7503 2024-2025Human Communication Means (6) ❖Storage Technologies ❖Storage Media ❖Magnetic Tape ❖Digital Storage ❖Allows storage, preservation and distribution of sound, video, etc. ❖❖Any invariance across time, space, and addressee? Distribution Technologies ❖Sending physical copy ❖Copy by wire ❖Internet distribution ❖Any invariance across time, space, and addressee?Images acquired from COMP7503 2024-2025Communications- related Inventions COMP7503 2024-2025Inventions Resulting Applicatoin Invariance Spoken Languages Written Languages Paper Print Telegraph Telephone Radio Television Movies Recording media Digital media InternetNatural communication None Symbolic record oflanguage Time Portability Time, space Mass distribution Time, space, addressee Remote narrow communication Space Remote analog communication Space Broadcasting ofsound Space, addressee Broadcasting of sight and sound Space, addressee Recording of sight and sound Time, space, addressee Recording Time, space, addressee Machine enhancement and processing Time, space, addressee Personalized reception Time, space, (addressee)TheroleofComputer inMultimedia during 1960s -era ❖During 1960s, computer isbulky ❖e.g.IBM1620 ❖No ALU, arithmetics processing is done based on lookup tables ❖Hardware ❖CPU a.k.a. CADET ❖Can’t Add, Doesn’t EvenTry ❖RAM: 60kbytes ❖I/O: Paper Tape reader and writer, or Punch Cards ❖Softwa re ❖Symbolic Programming System (assembly language) ❖FortranA 1960s- era Computer at NASA Image adopted from .1 in thetextbook COMP7503 2024-2025TheroleofComputer inMultimedia Before 1990s ❖Before 1990s, computing power, storage, bandwidth, and processing algorithms were not capable to deal with audio and video ❖Using multimedia in computer was under R&D phase ❖Hardware ❖CPU < Intel 80286, <6MHz ❖RAM < 640KB ❖Graphics Modes: 320x200 / 640x200, 16 colors ❖Storage: Floppy Drive, Harddisk <=20MB ❖Softwa re ❖OS:MS-DOS ❖Mostly text- based applications PC XT286 Image adopted from COMP7503 2024-2025TheroleofComputer inMultimedia Early 1990s ❖Early 1990s, high end PC’s are equipped with multimedia capabilities, like IBM PS/2 ❖Hardware ❖CPU Intel , 50- 60MHz ❖RAM <=64MB ❖Graphics Modes: 640x480, 1024x768, k colors ❖Storage: Floppy Drive, Harddisk <=400MB ❖May be equipped with dedicated Sound Card (e.g. Soundblaster) ❖Softwa re ❖OS:MS-DOS ❖GUI crafted byApplication ❖3D Gaming PossibleIBM PS/2 3D Games Wolfenstein on COMP7503 2024- PCTheroleofComputer inMultimedia Mid1990s ❖Around mid 1990s, more high end PC’s evolved, like Acer Presario 4122 ❖Hardware ❖CPU Intel Pentium 150MHz ❖RAM up to 128MB ❖Graphics Modes: up to 1280x1024 ❖Storage: Floppy Drive, 2.5GB harddrive, 8xCD-ROM ❖Integrated 16-bit Stereo Sound ❖Software ❖OS: Windows 95 ❖Audio Player, Video Player ❖CorelDraw 5 ❖Various kinds of gamesCompaq Presario 4122 Image adopted from of Experiential Environments (2) ❖Provide the same query and presentation spaces ❖Query and presentation spaces are the same (WYSIWYG) ❖Example(s) ❖Spreadsheet ❖Query :Change ofcertain data that is displayed inthecontext ofother data items ❖Presentation: Results in a new sheet showing new relationships COMP7503 2024-2025Properties of Experiential Environments (3) ❖Consider both the User State and Context ❖People operate best when they are in known contexts ❖Good experiential environment should promote perceptual analysis and exploration ❖Example(s) ❖Video Games effectively use audio, video and tactile media to create compelling interactive environments COMP7503 2024-2025Why Multimedia Computing? COMP7503 2024-2025❖Weneed agood experiential systems toimprove human -machine synergy forcurrent information environments ❖Experiential Systems are technically implemented as Multimedia Systems ❖Integrate, Process and Output data from different sensory modalitiesFormal Definition of Multimedia Computing ❖Consider a system equipped with multiple sensors working in a physical environment ❖Assume that S1, … , Sn are synchronised data streams from sensors ❖These data streams have K types of data in the form of ❖Image sequence, Audio stream, Motion detector, etc. ❖Let M1, … , Mn be the metadata for each stream ❖Annotations, Parameters, etc. ❖Let Fij be the j-th feature stream from Si M i COMP7503 2024-2025 Formal Definition ofMultimedia Computing (Cont’d) S1 M1 SiMi Mn……Feature Detector 1 Type 1{ Type i{ { Sn… … Type K…Feature Detector i Feature Detector n… …F11 F12 F1m Fi1 Fi2 Fip Fn1 Fn2 Fnq… } } } Integrate and ProcessMultimedia Computing Insights User 1 User’s feedback (query) as input data stream COMP7503 2024-2025❖Multimedia differs from monomedia fields like computer vision or audio processing in the way that ❖Multimedia is closely related to how humans experience the world ❖Partial information from multiple media sources is correlated and combined to get complete information about the environment ❖Multimedia computing and communication is fundamentally about combining information from multiple sources in the context of the problem being solvedExample of Multimedia Computing Application COMP7503 2024-2025Input: Metadata Output: Insight (route of trip) Image acquired from -glossary/what -is-metadata /

COMP7503 Multimedia Technologies Elements of Multimedia ComputingDr. Bill LUO COMP7503 2024 -2025 1Objective of Multimedia System COMP7503 2024 -2025 2❖Objective ❖Process different types ofdata streams simultaneously as onecorrelated setofstreams that represent information andknowledge ofinterest forsolving aproblem Challenge ❖Discover correlations that exist in the set of multimedia data ❖Combine partial information from disparate sources to build holistic information in a given context -republic.com/en/news/les -puzzles -1-000-pieces -sont -arrives -❖Experience andInforma tion COMP7503 2024 -2025 3❖ Information is an efficient but abstract communication of experience ❖We gain knowledge through the set of experiences that make up our lives and communicate information about those experiences❖Experience is the direct observation of or participation in events as a basis of knowledge (Webster’s dictionary ) ❖We experience the world we live in, and learn about the world and accumulate and aggregate our experiences in the form of knowledge Communication is the process of sharing experiences and information with others ❖Allows sharing of experiences with people who may be spatially and temporally separatedCommunication COMP7503 2024 -2025 4❖We gain knowledge through the set of experiences that make up our lives and communicate information about those experiences ❖But we don’t communicate the experiences themselves ❖During the communication process, we lose the vital element of the act of experiencing ❖Example ❖Information: “Wearing a mask is a powerful weapon under epidemic situation ” ❖For a man , he/she won’t know how important it could be, unless he/she has been there.Communication (Cont’d) COMP7503 2024 -2025 5 Communication (Cont’d) COMP7503 2024 -2025 6 April 16, 2003 –a main sitting on an empty MTR train. Photo: Peter Parks/AFPCommunication (Cont’d) COMP7503 2024 -2025 7❖Implications ❖For a man , he/she won’t know how important it could be, unless he/she has been there. ❖Effective communication of information requires the support by relevant experiences What can we do if we can not share the experience?Communication Process COMP7503 2024 -2025 8❖Dictionary is one of the most important elements forcommunication ❖Dictionary is a shared and agreed upon collection of symbols (words) and what these symbols mean ❖In language, dictionary is an exhaustive collection of a selection of words of a language ❖ ❖In Computer Science, dictionaries are extended to use lists of codes, terms, and keys for use by computer programmes ❖Compress algorithms also employ dictionary concept to achieve compression In multimedia, instead of traditional words or codes, the basic elements in the corresponding dictionary is called percept ❖Visual dictionaries ❖Curves, shapes, lines, etc.Objects andEvents COMP7503 2024 -2025 9❖ ❖Objects and events should be used to model a dynamic world ❖Objects are good at capturing static components of theworld ❖Events represent changes in relationships among objects Multimedia, in particular audio and video, are fundamentally dynamic in nature ❖They capture signals that represent some attributes of the world as a function oftimePerception COMP7503 2024 -2025 10❖Perception is the process of understanding sensory signals to recover useful information ❖The field attracted attention from psychologists and researchers in artificial intelligence ❖Understanding perceptual processes remains a difficult problem in many disciplines like psychology, neuro -physics, computer science, and computer vision ❖Understanding of sensory information is an important step in many multimedia systems ❖A perception system takes sensory signals as input and generates the information that is relevant in its application context asoutputPerception (Cont’d) ❖Without any knowledge, the system cannot produce any information ❖Perception sometimes is considered as a controlled hallucination process ❖Given a signal, the system starts simulating and creates multiple hypotheses to find the best supported hypotheses to recover information Perception Signal/Data COMP7503 2024 -2025 11Existing Knowledge InformationPerception (Cont’d) ❖What can you see from the image below? COMP7503 2024 -2025 12Perception (Cont’d) ❖What can you see from the image below? COMP7503 2024 -2025 13Perception (Cont’d) ❖What can you see from the image below? Dalmatian Dog COMP7503 2024 -2025 14Perception (Cont’d) ❖What can you see from the image below? Dalmatian Dog COMP7503 2024 -2025 15Perceptual Cycle ❖Neisser’s Perceptual Cycle ❖An agent is continuously interacting with the environment using its sensory mechanisms ❖Build the model (a.k.a. schema ) of the environment ❖The system then decides what is further required to complete the task and how that information could be acquired Environment (available information) Exploration Schema DirectsSamples Modifies COMP7503 2024 -2025 16AgentWine ??????? or MTR Mei Foo Station COMP7503 2024 -2025 17Wine Soy Sauce or MTR Mei Foo Station COMP7503 2024 -2025 18Problems toSolve COMP7503 2024 -2025 19❖How do we represent data in the most compact form for communication and storage? ❖How do we present vast amount of data to user’s computing environment to communicate intended information? ❖What system issues must be solved to deal with disparate types of data and how the system handles them? ❖How do we combine these data streams to obtain the information that is essential for solving the problem at hand?Semantic Gap COMP7503 2024 -2025 20❖Data in computing systems ❖Represented as bits and bytes ❖More sophisticated representations ❖Lists ❖Images ❖Aud io ❖Video ❖Users, on the other hand, define their applications based on abstractions of objects and events ❖There is a fundamental gap between different abstractions employed by computing systems and those by the usersSemantic Gap(Cont’d) ❖The semantic gap is the difference between the information that a machine can extract from (perceptual) data and the interpretation that a user in a given situation has for the same data ❖Multimedia computing needs to bridge the semantic gap between human and computer COMP7503 2024 -2025 21Context andContent ❖Context is the interrelated conditions in which some data (the content) is acquired ❖Exchangeable image file format (Exif) metadata for the photo COMP7503 2024 -2025 22Context and Content(Cont’d) COMP7503 2024 -2025 23❖ ❖ ❖Metadata can be regarded as “Data about data” Metadata can provide context parameters that may help in understanding the content The importance of context is becoming increasing clear ❖Content and context should be combined and should be viewed as all information that must be used for understanding multimediaComponents of a Multimedia System COMP7503 2024 -2025 24❖Capture Devices ❖Video Camera ❖Video Recorder ❖Microphone ❖Keyboards ❖Mouse ❖Touch Screen ❖Pressure sensor ❖etc.❖ ❖Storage Devices ❖Hard disks, Flash Drive, Optical Disc, etc. Communication Networks ❖Local network, 4G/5G connection, bluetooth, RF,etc. ❖ ❖Computer Systems ❖Workstations, Servers, DSP Output Devices ❖LCD Display, Speakers, Vibration Motor, Haptic Feedback, etc.

COMP7503 Multimedia Technologies Example ApplicationsDr. BillLuo COMP7503 2024-2025 1Audio Processing COMP7503 2024-2025 2❖ ❖ ❖ ❖❖Noise Cancellation Digital Audio Equalizer Speech -To-Text Shazam Audio CompressionNoise Cancellation ❖Noise Cancellation Headphone COMP7503 2024-2025 3Noise Cancellation ❖Noise Cancellation in Cars COMP7503 2024-2025 4Noise Cancellation ❖Noise Cancellation Microphone Mic 2 Mic 1Ambient NoiseSpeec h Signal Subtract❖iPhone s has 2microphones Amplify COMP7503 2024-2025 5Noise Cancellation (Cont’d) ❖iPhone 5s/+ have 3 microphones, why? iPhone 6s/6s+ got one more mic COMP7503 2024-2025 6Voice assistants being hampered bypoor microphone tech ❖Since the launch of the iPhone 5 in 2012, microphone technology hasn't advanced significantly, IHS Markit analyst Marwan Boustany explained to Bloomberg. As a result, mics still have difficulty picking up distant voices and filtering out background sounds. Even without these issues, keeping a mic on all the time —needed for voice triggers like "Hey Siri" —can sometimes consume too much battery life. ❖Companies likeApple aresaid tobelooking forbetter mics from suppliers tofixthese problems, aswell asachieve ahigher acoustic overload point less likely tobetripped . Atthesame time, size and power consumption need tobekept under control . ❖Partly to compensate for poor pickup, device makers have gradually been adding more microphones in recent years. While the first iPhone had a single mic, there are three in the iPhone 6, and four in the iPhone 6s. The Amazon Echo has seven, being a device that needs to hear users from anywhere in a room. ❖Some possible solutions may include mics with built -in audio processing algorithms, or ones using piezoelectric technology. ❖Microphones could become extremely important to Apple if the company is indeed developing an Echo competitor, whether in the form of a standalone product or an upgraded Apple TV. It's unknown how many mics the "iPhone 7" might be equipped with. Source: -siri-rivals -being -hampered -by-poor -microphone- tech COMP7503 2024-2025 7Digital Audio Equalizer ❖Allow fine tuning of the gain or attenuation of different frequency bands for the audio signal ❖Frequency bands cover the full audio range of 20Hz to20Khz ❖Number of frequency bands could be more than 200 bands for high end system ❖❖Gain/Attenuation ranges will vary from +/- 6 dB to +/ -24dB. Equalizer preset can be used instead of tuning individual gain ❖Bass Booster, Classical, Jazz, R&B, Rock, etc. Source: -amazon.com/images/I/61fGZ4pd55L.png COMP7503 2024-2025 8Speech -to-Text COMP7503 2024-2025 9❖ ❖ ❖ ❖ ❖A/D translates analog wave into digital data Noise reduction and signal normalisation is performed first Digitized audio data is divided into small segments (as short as a few hundredths/ thousandths of a second) Audio segments are then matches against known phonemes (~40 phonemes in English) Map phonemes to works through complex statistical modelShazam -Audio Search ❖Shazam is a service that takes short sample of music, and identifies thesong ❖Use time-frequency graph (spectrogram) to work out the “fingerprint” of each song in the catalog ❖User then take 10 second sample of audio, work out the fingerprint, and Shazam service would match this against its catalog ofmusic ❖ ❖Matching is based on spectrogram peaks ❖Robust to noise, room reverb, equalisation, overlapping sounds Original paper can be found at ❖ Audio ❖Frequency Masking ❖The ability of a loud sound in one frequency band to hide a softer sound in another frequency band that would have been audible in the absence of the loud sound. ❖Temporal Masking ❖After a loud sound stops, a quiet sound will be inaudible for a short period of time because the ear turns down its gain when it starts and it takes a finite time to turn it up again. 19COMP7503 2024-2025MPEG Audio (Cont’d) COMP7503 2024-2025 20❖ ❖Three layers, each layer applies additional optimisations ❖Layer 1: Basic Scheme (used in DCC digital tapes) ❖Layer 2: Advanced BitAllocation ❖Layer 3: Adds hybrid filters MP3 encoding ❖Process samples in groups of 1152 (about 26 msec worth) ❖Divide the samples into 32 frequency bands ❖Determine masked frequencies (using some psychoacoustic model) ❖Use variable length coding to explore coding redundanciesAudio Processing COMP7503 2024-2025 21❖ ❖ ❖ ❖❖Noise Cancellation Digital Audio Equalizer Speech -To-Text Shazam Audio CompressionVideo Processing and Analysis COMP7503 2024-2025 22❖ ❖ ❖ ❖ ❖ ❖ ❖ ❖Image/ Video Enhancement Video Stabilization Motion Deblurring Face Detection Head Tracker 3D Face Reconstruction Human Gait Recognition Video RetrievalImage Enhancement ❖Night Mode Enhancement ❖by capturing multiple frames shot at various shutter speeds. These frames are then intelligently combined in order to both recover shadow detail and prevent motion blur if your subject happens to bemoving. Ref: Apple Special Event Sep 2019 COMP7503 2024-2025 23Image Enhancement ❖Deep Fusion ❖Each time you press the shutter on the new 11 Pro or 11 Pro Max, the camera captures nine total images: four short images, one long exposure, and four secondary images. These images are then intelligently blended together “to optimize for detail and low noise.” COMP7503 2024-2025 Ref: Apple Special Event Sep 2019 24Video Enhancement ❖Retinex COMP7503 2024-2025 25Video Stabilization ❖ COMP7503 2024-2025 26Stabilises shaky video footage Deblurring ❖Motion Blurring Effect Due to global camera motion COMP7503 2024-2025 27Due to fast moving objectsMotion Deblurring (Cont’d) ❖ ❖Different portions from the same object might have different velocities w.r.t. camera Cannot assume single global motion for deblurring COMP7503 2024-2025 28Motion Deblurring (Cont’d) ❖Deblurring Processing Workflow 52C COMP7503 2024-2025 29Motion Deblurring (Cont’d) ❖Deblurring Example 53 COMP7503 2024-2025 30COM 2019Face Detection (Feature based) ❖ ❖Haar feature based ❖Proposed by Viola and Jones ❖Can be computed efficiently Adaboost Classifier ❖Combine ‘weak’ features into a ‘stronger’ classifier ❖For each image, the face detector scans each location to look for presence offace ❖May need to try scaling up or down for different face sizes Haar Features Adaboost ClassifierCOMP7503 2024-2025 31Head Tracker (model based) COMP7503 2024-2025 32Head Tracker (Cont’d) ❖ ❖Side View Modelling View from Left to Right COMP7503 2024-2025 33Head Tracker (Cont’d) People Counting by TripwireHead Captured when passing the tripwire COMP7503 2024-2025 343D FaceReconstruction (Model based Cont’d) ❖Face recognition works well with frontal face images only COMP7503 2024-2025 353D Face Reconstruction (Cont’d) ❖But face recognition does not work well with profile face image 30 COMP7503 2024-2025 363D Face Reconstruction (Cont’d) Facial Feature Detection 3DModeling Pose Estimation Texture Map Rebuilding COMP7503 2024-2025 373D Face Reconstruction (Cont’d) ❖ ❖ ❖Using active appearance model (AAM) to locate detailed facial features Using three distinct appearance models, which are frontal, half -profile and profile Individual models are then applied to match the input face image and search for the best fit. The one with minimum residual error isadopted Frontal COMP7503 2024-2025 38Half -Profile Profile3D Face Reconstruction (Cont’d) COMP7503 2024-2025 39❖Employed 3D face mesh has 6292 vertices and 6152 facet ❖A texture map contains all thetexture of a human face ❖Texture coordinates determine how the texture is mapped from the 2D texture map onto the final 3Dmesh3D Face Reconstruction (Cont’d) ❖Face Model Modification COMP7503 2024-2025 403D Face Reconstruction (Cont’d) ❖Pose Estimation COMP7503 2024-2025 413D Face Reconstruction (Cont’d) ❖Texture Map Rebuilding Rendering Fill & Interpolation Postprocessing CO42COMP7503 2024-20253D Face Reconstruction (Cont’d) ❖Reconstruction Examples COMP7503 2024-2025 433D Face Reconstruction (Cont’d) ❖Transform non-frontal face to frontal face image Face Reconstruction COMP7503 2024-2025 443D Face Reconstruction (Cont’d) ❖The profile face can be recognised properly after it is transformed to frontal face by Face Reconstruction 39 COMP7503 2024-2025 453D Face Reconstruction (Cont’d) Gallery image with frontal faces COMP7503 2024-2025Non- frontal faces Similarity to gallery image 0.7714 0.3341 0.1151 0.0000 Similarity to gallery image 0.8720 0.5469 0.6988 0.8474 46Face Recognition Limitation •Researchers from Carnegie Mellon University have shown that specially designed spectacle frames can fool even state- of-the-art facial recognition software. Not only can the glasses make the wearer essentially disappear to such automated systems, it can even trick them into thinking you’re someone else. •Reference: / //facial - recognition -glasses- trick- impersonate -fool COMP7503 2024-2025 47Face Recognition Limitation •Researchers at the University of York showed, even the smallest change to someone's facial appearance, like wearing glasses, can shift our ability to identify them if we don't know them •Reference: //31/ health/superman- glasses - disguise -facial -recognition/ index.html COMP7503 2024-2025 48iPhone X~14FaceID COMP7503 2024-2025 49Human GaitRecognition ❖Gait is one of the most important biometrics. COMP7503 2024-2025 50 KGB trained gun style -like-the-kgb- get-vladamir- putins -gunslinger -gaitHuman Gait Recognition (Cont’d) ❖Taking gait as a biometric offers potential for identifying human at low resolution when the subject only consists of few image pixels COMP7503 2024-2025 51Human Gait Recognition (Cont’d) ❖Sample of code words 47 COMP7503 2024-2025 52Human Gait Recognition (Cont’d) ❖Bag-of-Gait Representation Input Silhouette 0 1 0 0 0 0Feature vector 010000 COMP7503 2024-2025 53Human Gait Recognition (Cont’d) ❖Gait Tracking Demo 49 COMP7503 2024-2025 54Video Retrieval ❖Overview Tracking Feature Extraction For Each Tracked ObjectDBVideo Source Compressed Video StreamsH.264/MPEG -4 Encoder Motion Vectors and Background Model Moving Object Segmen tation Retrieval Request COMP7503 2024-2025 55Video Retrieval (Cont’d) ❖Video Indexing INSIGHTS ANALYTICS METADATA VIDEO DATA COMP7503 2024-2025 56Video Retrieval (Cont’d) ❖Color and Edge Features blue purple red yellow green 0°45°90°135°Databa se COMP7503 2024-2025 57Video Retrieval (Cont’d) ❖Retrieval Examples COMP7503 2024-2025 58Video Retrieval (Cont’d) ❖Retrieval Examples COMP7503 2024-2025 59Dialogue in Metaverse COMP7503 2024-2025 60 Lex FridmanMark Zuckerberg Ref: Information, Knowledge, Wisdom (DIKW) Pyramid Ref: //knowledgehub/fundamentals/dikw -pyramid/ COMP7503 2024-2025 61DataAcquisitionsDataCorrelations DataCleansing DataTaggingPrescriptive Analytics Predictive Analytics Insights, Decision Making Reporting, Dashboards Business Analytics DataTaggingAnalytics in Daily Lives Ref: //en/company/innovation/customer -solutions/data -analytics.html 61COMP7503 2024-2025 62Smart City COMP7503 2024-2025 63 -mario

COMP7503 Multimedia Technologies Media Sensor and Reproduction -Part1Dr. Bill LUO COMP7503 2024-2025 1Introduction toSensors COMP7503 2024-2025❖Information about the environment is always obtained though sensors ❖A sensor is a device that measures a physical quantity and converts it into a signal that a human or a machine can use ❖The most important sensors for multimedia computing ❖Sound sensor ❖Light sensor 2Types ofSensors COMP7503 2024-2025❖ ❖ ❖Sensors that imitate human sensors Sensors that do not imitate human sensors ❖Measure aspects oflife Sensors that measure physical facts about the environment 3Human Sensors COMP7503 2024-2025❖Correspond to 5senses ❖Touch ❖Sight ❖Hearing ❖Smell ❖Taste 4Touch Sensor ❖ ❖Sense organ fortouch ❖Skin -The largest organ on human body Sensors that imitate human skin ❖❖Tactile sensors -Translate pressure into signals Pressure Sensor, Resistive Sensor and Capacitive Sensor Source: Resistive%20Touch%20Screen%20Sensor 5COMP7503 2024-2025Touch Sensor 3D Touch Technology COMP7503 2024-2025 6Touch Sensor ❖Feedback is provided byactuators ❖Vibration module or Taptic Engine COMP7503 2024-2025 7Sight Sensor ❖Sense organ forsight ❖Eyes ❖A big portion of human brain isdevoted to the sight sensation ❖Sight sensors are collectively called cameras ❖Sensing elements ❖Based on electrical photovoltaic principles ❖Charge- Coupled Device (CCD) ❖CMOS Sensor ❖The more sensor elements, the higher would be the resolution References: 1. the-hole -in-the-hand- a-look -into-the- neural -software -of-the-visual -brain/ 2. SustainableEngineering/ RenewableEnergyEngineering/ SolarEnergyEngineering/Photovoltaics/ tabid/3890/Default.aspx COMP7503 2024-2025 8Spot thedifferences Source: 9COMP7503 2024-2025Spot thedifferences Source: 10COMP7503 2024-2025Spot thedifferences Which brain areas you used just now ❖Red region : You have to identify the objects that you see: this involves your occipital lobes. ❖Green region : You have to analyzed the spatial relationships between the objects that you see: this involves your occipital and parietal lobes. ❖Blue region : You have to remember what you see in one picture and compare it to what you see in the other picture, that is you have to use your short -term memory: this involves your frontal and parietal lobes. ❖Blue region : You have to mark down the locations where you see a difference: this involves mostly your frontal lobes (for the movement) brain -teasers- spot -the-difference/ 11COMP7503 2024-2025Sight Sensor ❖Binocular Vision Ref: 12COMP7503 2024-2025How to seea Stereogram ❖Bring the stereogram image really close to your eyes (until you touch it with your nose). At this distance your eyes cannot focus on the image and they look somewhere behind the image. Now, slowly push the image away from you, while trying to keep the eyes off focus. At some point you will see the hidden image. ❖Another method is to take an object and put it behind the image (about half of meter behind it). Now, focus on the object behind the image while keeping the eyes looking at theimage. COMP7503 2024-2025 1314COMP7503 2024-2025Ref: -microphones/ -basics- transducers- polar -patterns -frequency -response/Dynamic Microphones Condenser Microphones COMP7503 2024-2025 16Hearing Sensor COMP7503 2024-2025 17 2021 Best Selling Dynamic 2021 Best Selling Condenser Smell &Taste Sensors ❖ ❖Sense organs for smell and taste ❖Nose & Tongue Chemistry and physics of smell and taste are not well understood yet ❖No artificial noses and tongues have been properly produced yet ❖For this course, we shall focus primarily the sense of sight and hearing Reference: -BoxEx -18-01-B-0.jpg COMP7503 2024-2025 19Sensors about Life ❖Sensors used to measure life vitals or health statistics ❖Weight sensor ❖To measure body weight ❖Together with height, gender, and age, wecanalso estimate theBMI values ❖ ❖❖Temperature sensor ❖To measure body temperature Noise sensor ❖To detect loud environment Heart rate sensor❖ ❖❖To measure heart rate ECG sensor ❖To detect Sinus Rhythm Movement sensor ❖To measure motion/ acceleration magnitudes ❖Together with heart rate, estimation of burned calories ispossible ❖These sensors are becoming more abundantly available, so so they are likely to be very important for future multimedia computing COMP7503 2024-2025 20Sensors about Life COMP7503 2024-2025 21 Care -on-Call Service for senior citizen Environment Sensors ❖ ❖ ❖❖ ❖❖Accelerometers ❖To sense device orientation ❖Can provide data forpedometer Gyroscopes ❖To sense device rotation motion GPS receiver ❖For location sensing Radiation sensor ❖For sensing radiation Altimeter sensor ❖For sensing altitude Sometimes, environment sensors is more reliable than human perception ❖A person may suffer from spatial disorientation Reference: -of- aeronautical -knowledge/ch -7-flight -instruments/gyroscopic - COMP7503 2024-2025flight -instruments22Sensors Limitations ❖Limited dynamic range ❖Possible intensities of the input signal lie within a certain interval some shadow and highlight detail is lostcapture shadow detail COMP7503 2024-2025 23at the expense of highlight detail capture highlight detail at the expense of shadow detailWashed outimageReference: imaging/dynamic -rangeSensors Limitations (Cont’d) ❖Offset orbias ❖The sensor output is not zero when the input is zero ❖Nonlinearity ❖A linear increase in input signal does not result in a linear increase in the sensor output ❖Some sensor can be tuned to behave linearly inside an operational range ❖Hysteresis ❖Sensor Response Deviation over time A general example for the hysteresis concept. The bottom curve is the sensor input, and the upper curve is the output. As the arrows show, the curve behaves differently when the measured entity decreases compared to when it increases. COMP7503 2024-2025 24How to build a LEGO House COMP7503 2024-2025 25 -a-LEGO -HouseSensor Output Analyze t COMP7503 2024-2025 26 Sensor Output Analyze t COMP7503 2024-2025 27 Sinusoid Signal ❖A sinusoidal signal can be regarded as the projection on a plane of the path of a point moving around a circle at uniform speed A -A A -AA -A t COMP7503 2024-2025 28Sinusoid Signal (Cont’d) ❖A sinusoidal signal can be regarded as the projection on a plane of the path of a point moving around a circle at uniform speed A -A A -AA -A t COMP7503 2024-2025 29Sinusoid Signal (Cont’d) ❖A sinusoidal signal can be regarded as the projection on a plane of the path of a point moving around a circle at uniform speed A -A A -AA -A t COMP7503 2024-2025 30Digitization ❖Sensor outputs continuous electrical signal (voltage/current), and these signal needs to be digitised for computer tooperate ❖Digitization is a process for converting continuous signal into a discrete set of its samples ❖It happens in both the time and signal values ❖In a sampling step, the analog signal is captured at regular time intervals (the frequency of taking sample is called sampling rate) ❖Each reading is called a sample, and each sample is rounded to a fixed set of integer numbers ❖This is called quantization process ❖Digitized sample values are much less susceptible to distortions than analog signalsDigital representation of an analog signal. Both amplitude and time axis arediscretised. COMP7503 2024-2025 31 Digitization (Cont’d) COMP7503 2024-2025 32 Digitization (Cont’d) ❖ ❖The error the quantisation introduces is the quantization noise ❖It affects how accurately the amplitude can berepresented ❖Few bits ==> Coarse representation ❖More bits ==> Finer representation For audio and video, eight, sixteen, twenty -four and thirty - two bits per sample are used ❖The error the sampling rate introduces is the discretization error ❖It determines the maximum frequency that can be represented in the signal ❖The maximum frequency is called Nyqvist Frequency ,which is half the sampling frequency (rate) fs ❖If a function x(t) contains no frequencies higher than B hertz, x(t) is completely determined by its ordinates at a series of points spaced 1/(2B ) seconds apart Digital representation of an analog signal. Both amplitude and time axis arediscretised.tQuantization errors COMP7503 2024-2025 33Nyquist Frequency ❖Necessity of fs >2B ❖Consider the sinusoid ❖ ❖ ❖which render samples that are indistinguishable from the samples of ❖ Three possible analog signals for the same sampling points With fs = 2B ( i.e. T = 1/(2 B)),the samples are given by 0 ❖ COMP7503 2024-2025 34Physical Properties ofSound ❖Sound ❖“Atravelling (longitudinal) wave which isanoscillation ofpressure transmitted through a solid, liquid, orgas, composed offrequencies within therange ofhearing and ofalevel sufficiently strong tobeheard, orthesensation stimulated inorgans ofhearing bysuch vibrations .” ❖ ❖ ❖Transmission ❖Propagate in physical medium by mechanical oscillations/vibrations ❖Cannot travel through vacuum Speed of sound is ❖Mach 1 or 343 ms-1 in dry air at 20oC ❖1,482 ms-1 in water at20oC Human hearing frequency range ❖12Hz ~ 22kHz Reference: -waves.htmlCOMP7503 2024-2025 35Sound Pressure Level (SPL) ❖Sound Pressure Level (SPL), or L p , is defined as ❖ ❖Measurement on logarithmic scale using decibels ❖ ❖where p is the RMS sound pressure and pref is a reference sound pressure Typical reference sound pressure for silence is 20 µPa in air and 1 µPa in water ❖prefis often considered as the threshold of human hearing (roughly the sound of a mosquito flying 3 maway) ❖ ❖The sound pressure perceived by human ear is non- linear Human ear, does not react equally to all frequencies ❖The same sound pressure at a different frequency will be perceived as a different volume level COMP7503 2024-2025 36A-weighting scheme ❖Sound -pressure measurements are often frequency weighted to match perception. A-weighting is applied to instrument -measured sound levels in an effort to account for the relative loudness perceived by the human ear, as the ear is less sensitive to low audio frequencies. It is employed by arithmetically adding a table of values, listed by octave or third -octave bands, to the measured sound pressure levels in dB. Sound- pressure levels weighted by A -weighting scheme are usually labeled as dBA ordB(A) Ref: comparison -chart -decibels- dba/ 37COMP7503 2024-2025Undesirable Sound Effects COMP7503 2024-2025 38❖Echo, Reverberation, and Interference make automatic processing and recognition of sounds difficult ❖Echo ❖Echo is a reflection of sound, perceived by the listener some time after the original ❖e.g. Echoes produced by the bottom of a water well, or by mountain enclosures ❖Echoes are present in every environment because most materials easily reflect sounds ❖Human ear cannot distinguish between the echo and the original sound if the delay is less than secs ❖But a machine -based system trying to recognise sound might still have an issue with these echoesReverberation ❖Reverberation is created when a sound is produced in an enclosed space ❖Numerous echoes build up and slowly decay as the environment absorbs the sound ❖Multimedia content analysis techniques often suffer from not accounting for reverberation, even when it is inaudible Reference: COMP7503 2024-2025 39Interference ❖ ❖Interference is the superposition of two or more waves that results in a new wave pattern Destructive interference can be useful for certain scenario, can you think of any? COMP7503 2024-2025 40Primitive form ofRecording andReproduction ofSound ❖Sound pressure variation are captured and mechanically engraved on the outside surface of a strip of tinfoil wrapped around a rotating cylinder ❖To playback, a needle ran along the cylinder, to convert mechanical engravings into sound saves that would be mechanically amplified Reference: COMP7503 2024-2025 41Primitive form ofRecording andReproduction ofSound (Cont’d) COMP7503 2024-2025 42 Eighteen- note music boxMicrophones ❖Microphone is an acoustic sensor that converts sound into an electrical signal ❖A membrane is exposed to sound pressure, which create electrical properties variations (e.g. resistance, voltage, capacitance, etc.) ❖Directionali ty ❖This is an important characteristic that indicates the sensitivity to pressure waves arriving at different angles ❖It is usually represented by a polar pattern, visualising the locations that produce the same signal level output in the microphone if a constant sound -pressure level is generated ❖In real world, polar patterns are a function of frequency COMP7503 2024-2025 43Polar Patterns COMP7503 2024-2025 44 Cardioid ref: -patternsPolar Patterns ❖Omnidirectional ❖Omni mics are equally sensitive to sound arriving from allangles ❖❖Good for studio environments without noise Bad in rejecting background noise ❖ ❖Cardioid ❖Most commonly used directional polar pattern ❖Maximum rejection at 180 degrees off-axis Supercardioid ❖❖Narrower front pick angles than cardioid Maximum rejection at 120 degrees ❖Shotgun ❖Highly Directional Microphones Omnidirectional COMP7503 2024-2025 45Cardioid Supercardioid Shotg unMic Frequency Response ❖The electrical voltage output from microphone should be ideally directly proportional to the sound pressure ❖In reality, sound recording has a linear area for certain sound pressure levels and frequency ranges only Figure adopted from COMP7503 2024-2025 46Microphone Output ❖Analog Signal Output ❖Microphone’s analog output is feed to an amplifier before it is digitised ❖If the amplifier gain is not set properly ❖Signal can be overdriven ❖Signal with high amplitude can be clipped out, leading to distortion ❖On the other hand ❖Signal with small amplitude can zero out after A/D conversion Image adopted from vs-distortion/ COMP7503 2024-2025 47Microphone Output (Cont’d) ❖AES 42 Digital Output ❖No interference problem ❖Configurable Settings ❖Polar -pattern ❖Low -cutfilter ❖40Hz, 80Hz, 120Hz ❖Gain ❖Unity to +63dB ❖Mute ❖For more details, please refer to ❖ w.taper ssection.com/ reference/ pdf/AES%20 - %20Digital%20Microphones%20 - %20AES42%20and%20all%20that.pdf Reference: /img/Linkgraphics/Solution- D\_E.pdf COMP7503 2024-2025 48Human Audible Frequency Range COMP7503 2024-2025 49❖ ❖Human Auditory System ❖Hearing range is from 20 Hz to 22 kHz Professional audio -recording equipment usually employs sampling frequencies of at least 44kHz ❖Though 48 kHz and 96 kHz might also be used, even if the human ear cannot perceive them ❖Digital filtering and machine learning, might use these higher frequencies