Al-Mustaqbal University

Biomedical Engineering Department

Class: 3rd

Subject: Rehabilitation Science and Engineering.

Lecturer: Mr. Mahir Rahman Al-Hajaj

2nd term – Lect. 6: Augmentative and Alternative Communication (AAC) System Engineering.

Email: mahir.rahman@uomus.edu.iq

Introduction to AAC Engineering

- **Defining the Domain:** Augmentative and Alternative Communication (AAC) encompasses the methods, devices, and strategies used to supplement or replace biological speech and writing for individuals with severe motor or neurological impairments.
- **Historical and Technological Convergence:** The modern AAC field was born from a "perfect storm" where the progressive deinstitutionalization of individuals with disabilities intersected directly with the advent and miniaturization of the personal computer.

Introduction to AAC Engineering

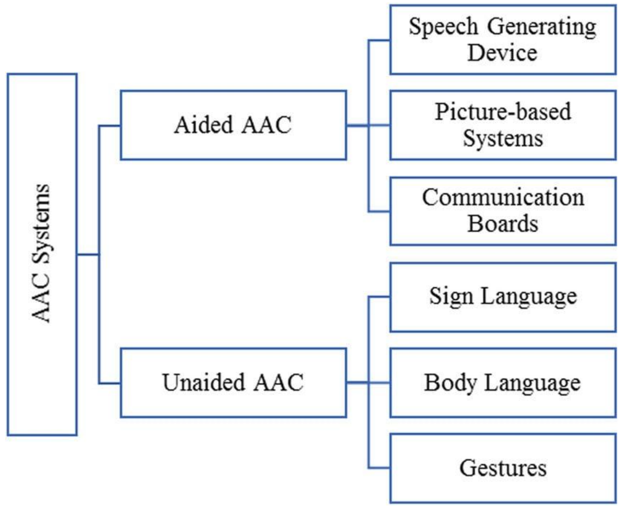
- **The Genesis of Rehabilitation Engineering:** Early pioneers (often engineering students) began inventing custom interfaces, such as the Tufts Interactive Communicator (TIC) and the Cybertype, to bypass damaged biological output pathways and augment physiological speech mechanisms.
- **Modern Engineering Focus:** Today, AAC is treated as a complex systems engineering problem focused on human-machine connectivity. The goal is to optimize signal transduction, minimize neurological processing loads, and balance design trade-offs to translate minimal physiological movement (or raw neural data) into complex linguistic output.

The Sender-Receiver Communication Model

- Quantitative Modeling of AAC: To engineer effective solutions, AAC relies heavily on the classical information-theory/sender-receiver model to break communication down into a discrete, measurable signal path. The Four System Components:
- 1. Operator: The human transferring intent.
- 2. Interface: The physical sensor or switch detecting the biological signal.
- 3. Controller: The microprocessor and algorithm translating raw input into selection.
- 4. Output: The final synthesized linguistic result.

The Sender-Receiver Communication Model

- Objective of the Model: By isolating these variables, engineers can perform quantitative assessments of communication efficiency.
- Key Optimization Metrics: The evaluation of the system's success is based on "switch selection efficiency" and "keystroke savings," which are critical metrics because every motor actuation costs a severely paralyzed patient massive physiological energy.



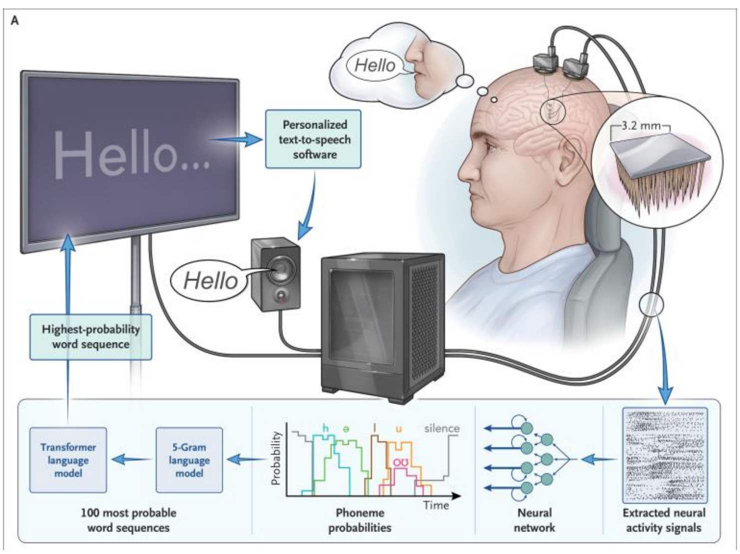
Evaluating Communication Throughput (WPM) & BCI Breakthroughs

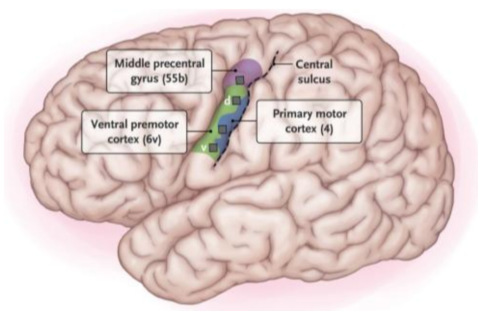
- **The Interface Bottleneck:** Natural human speech operates at 150-200 Words Per Minute (WPM). AAC systems relying on physical motor inputs are severely bottlenecked by interface bandwidth.
- A systematic review shows standard On-Screen Keyboards (OSK) yield ~15.4 WPM, while 1-switch scanning drops to ~1.7 WPM.
- **Optimizing Physical Encoding:** Adjusting the encoding algorithm can drastically improve data transmission rates even with minimal physical inputs; for example, utilizing a 2-switch Morse code system yields 12.5 WPM compared to 5.0 WPM for 1-switch Morse.

Evaluating Communication Throughput (WPM) & BCI Breakthroughs

- **Advanced Neural Interfacing (2023 Update):** The engineering landscape shifted dramatically when Intracortical Speech Neuroprostheses achieved 62 WPM with a 125,000-word vocabulary, bringing synthetic output much closer to natural conversational speeds.
- **Mechanism of Action:** By implanting microelectrode arrays directly into the motor cortex (area 6v) of an ALS patient, researchers used a Recurrent Neural Network (RNN) to decode attempted speech articulatory movements.
- This successfully bypassed the muscular bottleneck entirely by decoding spatially intermixed tuning to speech articulators.

Intracortical Speech Neuroprostheses





brain-to-text speech neuroprosthesis , DOI: 10.1056/NEJMoa2314132

The Cognitive-Physical Optimization Trade-off & LLMs

- **Paradox of Prediction Algorithms:** While standard word prediction algorithms can mathematically achieve 40-50% keystroke savings, they introduce a classic human-factors trade-off.
- **Cognitive Load vs. Mechanical Speed:** Dynamic prediction screens force the user to constantly break their motor plan to visually scan a changing list.
- Simulation and empirical data show that the time required to visually parse this dynamic data often negates the mechanical speed advantages, sometimes causing a three-fold speed decrease.
- Algorithmic Shift (2024 Update): Recent advancements have mitigated this by integrating Large Language Models (LLMs) like GPT-4 into AAC architectures, shifting prediction away from simple n-gram Markov chains to deep semantic intent analysis.

The Cognitive-Physical Optimization Trade-off & LLMs

- **Algorithmic Shift (2024 Update):** Recent advancements have mitigated this by integrating Large Language Models (LLMs) like GPT-4 into AAC architectures, shifting prediction away from simple n-gram Markov chains to deep semantic intent analysis.
- **Reducing Shannon Entropy:** Instead of predicting the next immediate word, LLMs analyze multimodal contexts (e.g., previous conversational turns) to generate entire high-probability sentences.
- This effectively reduces the Shannon entropy of the communication channel, allowing users (especially those using eye gaze) to communicate complex thoughts with a fraction of the physical actuations previously required.

Temporal Dynamics of AAC Use

- **Interaction Timeframes:** Time is a critical constraint in system design. AAC interaction is modeled in three distinct timeframes that dictate communication success.
- **Now-Time & Near-Time:** "Now-Time" consists of immediate, responsive expressions (biological speech, vocalizations, gestures). "Near-Time" involves rapid expression accomplished with well-designed AAC systems using one or two selections.
- **Delayed-Time Interactions:** "Delayed-Time" requires composing sentences letter-by-letter, taking many seconds or minutes.
- This delay results in frequent misunderstandings due to the communication partner's lapses in attention, forgetfulness, and distractibility.

Temporal Dynamics of AAC Use

- **Engineering Solutions:** Most current high-tech AAC devices structurally force users into Delayed-Time. To fix this design flaw, systems must integrate alternative views, expressive gestures, rapid access to pragmatically useful phrases, and shared visual interfaces to pull users back into Near-Time interaction.

Fundamentals of Speech Synthesis

- **Electronic Transduction of Human Voice:** Synthesizing speech requires mathematically modeling the human vocal tract. This began in 1939 with the VODER, which electronically transduced key presses into continuous human speech.
- **The DECtalk Breakthrough (1983):** Dennis Klatt's algorithm (commercialized as DECtalk) modeled the formants of human speech, allowing for the manipulation of fundamental frequencies to create distinct ages and genders, providing users with a voice that matched their identity.

Fundamentals of Speech Synthesis

- **Critical Acoustic Metrics:** In signal processing terms, the primary optimization vectors for synthetic speech are intelligibility (can the listener decode the phonemes?) and naturalness (does it possess human prosody?).
- **Modern Synthesis Generation:** Today's standard synthetic voices have at their foundation voice actors who record large scripts in controlled acoustic environments to represent a complete sampling of co-articulated speech sounds.

Advanced Voice Preservation Technologies (VITS)

- **The Need for Voice Banking:** Voice banking allows a person at risk of losing their speech (due to motor neuron disease, laryngectomy, etc.) to create a personal synthetic voice, preserving their unique acoustical fingerprint and emotional tone.
- **Traditional Constraints:** Historically, this process required massive datasets, forcing the speaker to record up to 1,600 sentences to capture all phonemic variations.
- **Deep Neural Networks (DNN):** The modern engineering paradigm has shifted to single-stage, end-to-end models like VITS (Variational Inference with Adversarial Learning).

Advanced Voice Preservation Technologies (VITS)

- **Zero-Shot Voice Cloning:** VITS utilizes a Conditional Variational Autoencoder augmented with normalizing flows.
- It mathematically models the uncertainty over latent variables, allowing it to generate raw audio waveforms directly from text.
- This enables "zero-shot" cloning, requiring mere seconds of a patient's historical audio to perfectly synthesize a highly natural voice, bypassing the older two-stage spectrogram-to-vocoder pipeline.

Alternative Access: The Control Interface

- **Access Engineering Framework:** When standard QWERTY keyboards fail, engineers design "Alternative Access" solutions, divided into three components.

1. **Input Device:** The physical hardware interface the user interacts with (e.g., mechanical switch, joystick, camera).
2. **Symbol Selection Set:** The software representation of choices (letters, words, pictures) presented in visual, auditory, or tactile formats.
3. **Selection Method:** The algorithm translating the hardware input into a software selection, categorized as Direct Access or Indirect Access.

Alternative Access: The Control Interface

- **Assessment Priority:** The engineering constraint is patient physiology. We must identify a movement that is volitional, repeatable, and non-fatiguing.
- If an interface triggers abnormal reflex patterns or causes a significant increase in abnormal tone (like total body extension), the signal-to-noise ratio drops drastically due to fatigue, rendering the device useless.

Transducing Direct Biological Movement

- **Direct Access Characteristics:** Direct access maps a user's biological movement 1:1 with cursor movement. It is fast and requires less cognitive skill, but demands high motor control fidelity.
- **Optical Tracking Technologies:** For severely paralyzed patients, we use optical tracking. Eye trackers utilize cameras emitting infrared signals that bounce off the retina; trigonometric algorithms calculate the exact gaze vector on the screen. Head trackers use similar computer vision techniques tracking a reflective dot on the forehead.

Transducing Direct Biological Movement

- **The "Midas Touch" Problem:** The fundamental engineering challenge here is determining intent, how do we mathematically differentiate between the user simply scanning the screen and the user intending to select a button?
- **Algorithmic Solutions:** Engineers solve this using programmable thresholds like specific "dwell times" (staring within a specific pixel radius for a set duration), blinking, or requiring activation of a secondary mechanical switch.

Indirect Access: Signal Transduction in Switches

- **The Role of Indirect Access:** When fine motor control is absent, we use indirect access via switches, reducing complex vocabulary selection down to a binary 1 or 0 action.
- **Mechanical Transducers:** Switch hardware varies widely based on patient capability, utilizing pressure difference, pneumatic (sip-and-puff), proximity, motion detection (mercury), or light sensitivity sensors.
- **Biological Transducers:** For profound paralysis, we detect physiological signals directly using myoelectric (EMG) muscle sensors or brain activity (EEG) sensors.
- **The Algorithmic Challenge:** The core engineering problem is system translation: efficiently converting a 1-bit, highly constrained input string into a rich, 150-WPM conversational output using encoding or scanning software.

Categorization of AAC Hardware Architecture

- **Low-Tech Systems:** Non-electronic boards (alphabet, symbols, objects) requiring zero power. Highly portable, reliable, and inexpensive, but produce no audible output to secure a partner's attention.
- **Mid-Tech Systems:** Electronic devices utilizing static overlays and simple microcontrollers with static EEPROM memory. They rely on "digitized" speech (directly recorded human audio clips played back upon activation).
- **High-Tech Systems:** Computerized devices with dynamic screens and real-time text-to-speech synthesized output algorithms. They provide massive storage, internet access, and environmental controls.

Interface Architecture: Visual Scene Displays (VSDs)

- **Cognitive Load of Traditional Grids:** The Graphical User Interface (GUI) architecture heavily dictates the cognitive load on the patient. Traditional displays arrange individual icons or words in an evenly spaced orthogonal grid.
- **The Abstraction Barrier:** Grids require intact cognitive-linguistic mapping skills, the patient must know that a generic, abstract icon of a cup means "drink." For patients with severe aphasia or traumatic brain injury, this level of abstraction fails.
- **Visual Scene Displays (VSDs):** Engineering research led to VSDs, which replace the grid with highly relevant, contextualized, holistic photographs of a specific event or activity (e.g., a photo of the patient's actual kitchen).

Clinical Systems Engineering Models

- **Beyond the Hardware:** Hardware implementation requires a systems engineering approach. Two complementary models currently represent best practice in AAC service delivery: The Feature Matching Model and the Participation Model.
- **The Feature Matching Model:** This acts as an operational matrix. Engineers quantify the user's specific capabilities (motor control, cognitive, linguistic, and sensory function) and mathematically map those variables to specific technology features (e.g., switch sensitivity, screen layout, output type).
- **The Participation Model:** Concurrently, this model evaluates the system within its operating environment. It identifies specific opportunity barriers (policy, facilitator skill, attitudes) and access barriers, ensuring the engineered device actually increases functional social participation in the real world.

The Future of AAC Engineering (Synthesizing Neural Code)

- **Bypassing the Musculoskeletal System:** The frontier of AAC is entirely engineering-driven, moving away from gross physical switches toward highly sensitive, wearable, and embeddable sensors.

- **Brain-Computer Interfaces (BCI):** AAC access methods increasingly rely on both non-invasive BCI (sensors placed in contact with the scalp) and invasive BCI (electrodes surgically implanted on the motor cortex).

- **Neural Decoding:** The connected computer translates raw brain activity into control signals. As demonstrated by recent studies, detailed articulatory representations of phonemes persist in the motor cortex years after paralysis. The ultimate engineering trajectory is scaling these arrays to create a closed-loop AI system capable of bypassing the physical body entirely—decoding cortical neural firing patterns and transforming those brain signals directly into high-fidelity human language.

Thank You
For Your Attention