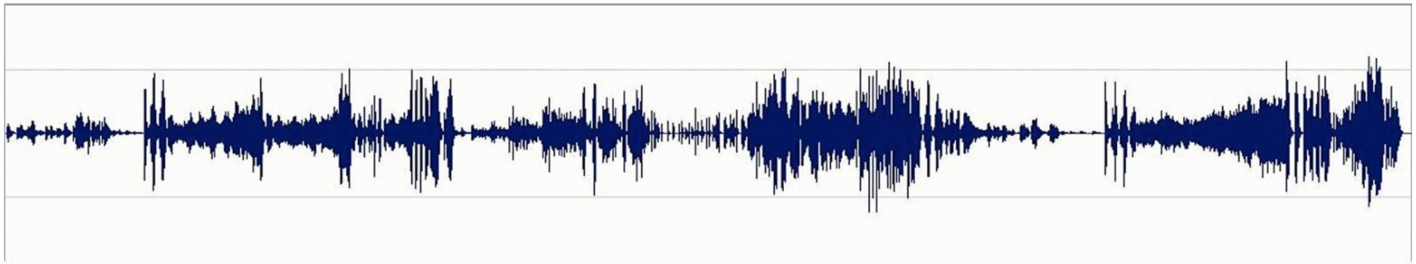
CHAPTER 2

Sound and Audio System

2.1. Basic Sound Concept: Representation and Format

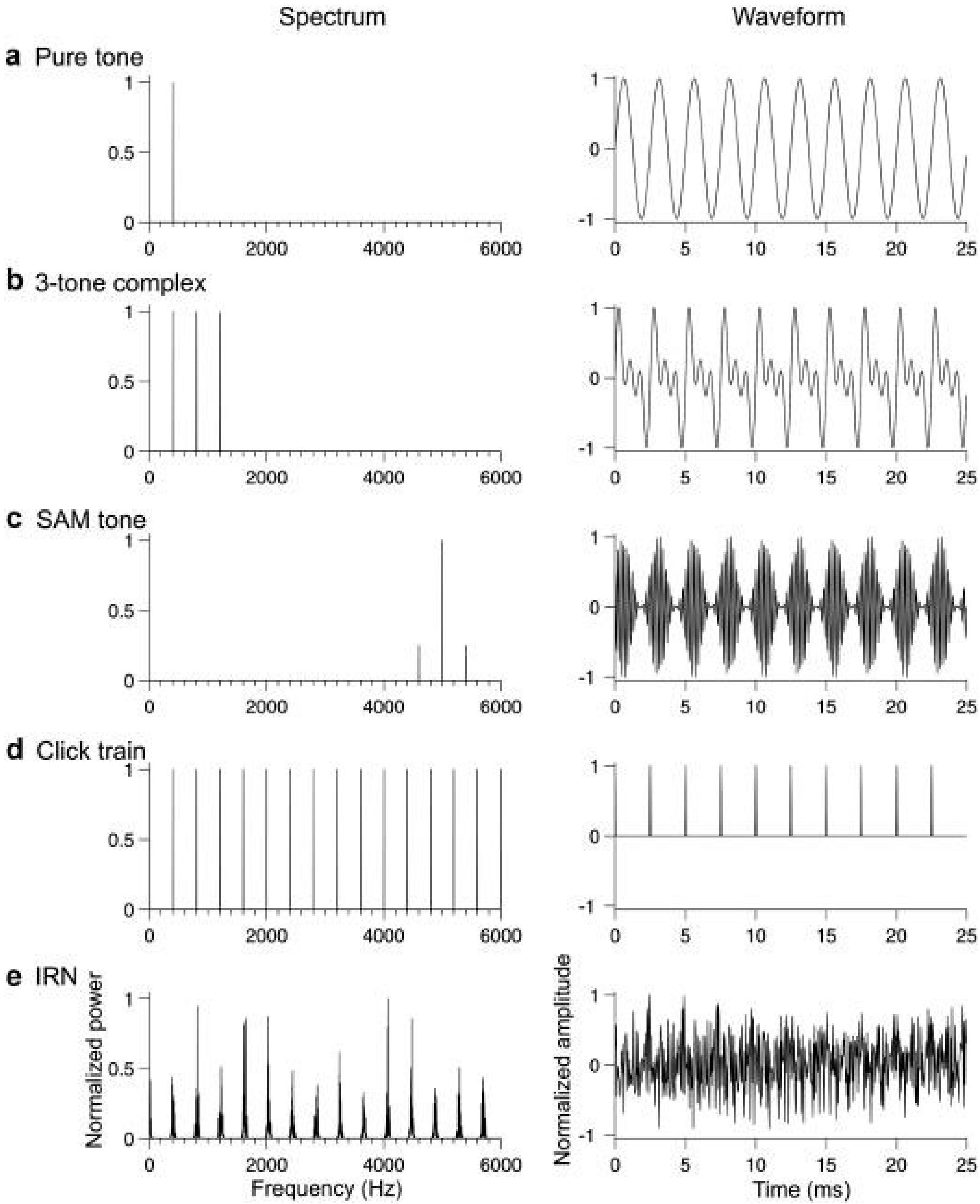
- ***Sound is the physical phenomenon produced by the vibration of matter*** such as violin string or block of wood.
- When a matter vibrates, pressure variations are created in the air surrounding it.
- This alteration of high and low pressure is propagated through the air in a wave like motion. When the wave reaches the human ear, a sound is heard.



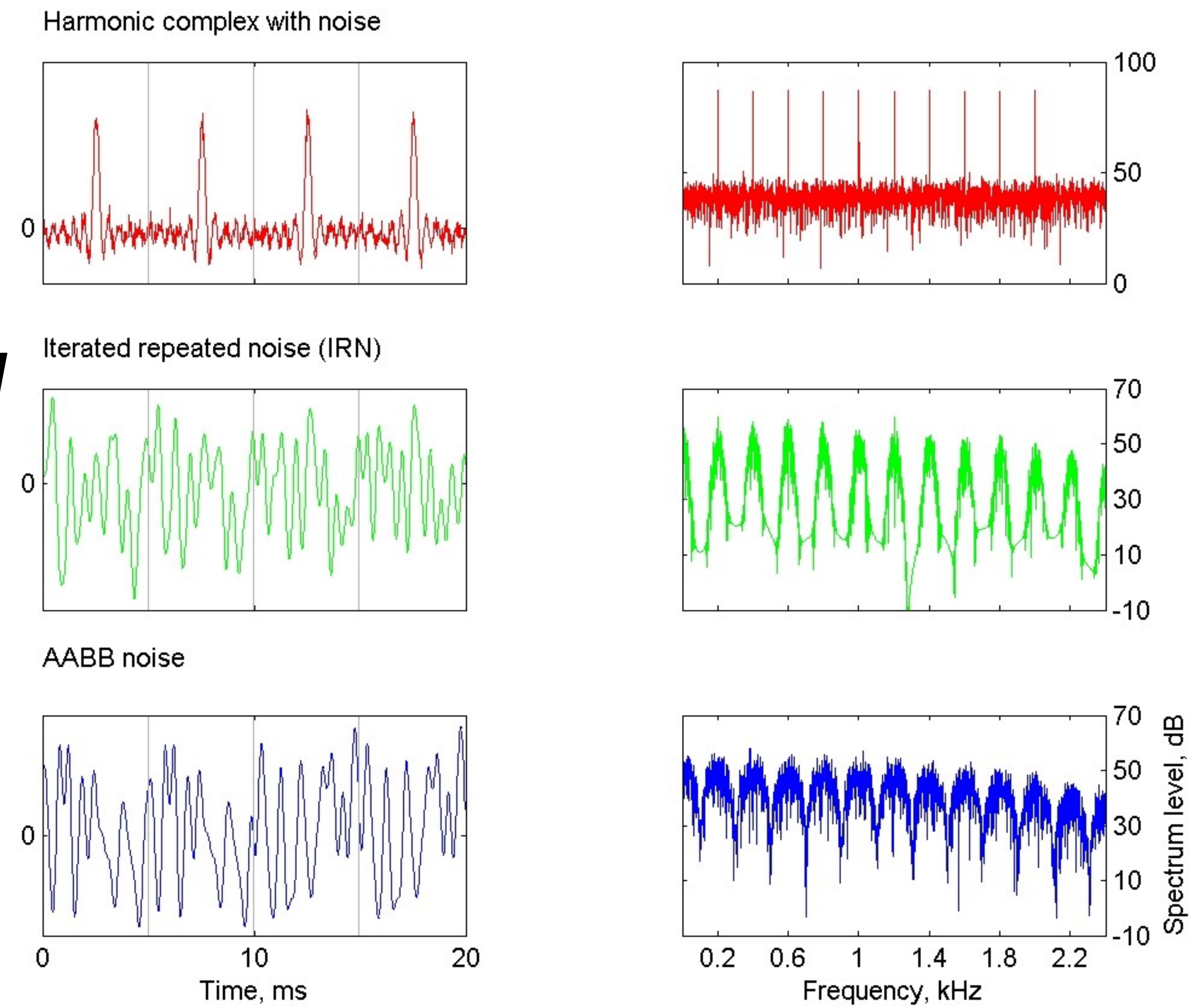
2.1. Basic Sound Concept: Representation and Format…

- The pattern of the oscillation is called a ***waveform***.
- The waveform repeats the same shape at regular intervals and this point is called a ***period***.
- ***Since sound wave forms occur naturally, Sound waves are never perfectly smooth or uniformly periodic.***
- ***Periodic sound****:* E.g. Musical instruments, vowel sounds, whistling wind, bird songs etc.
- ***Non-periodic sound****: E.g.* Unpitched percussion Instruments, coughs, sneezes, rushing water, Consonants,.

Periodic sound



Non Periodic sound



2.1. Basic Sound Concept: Representation and Format…

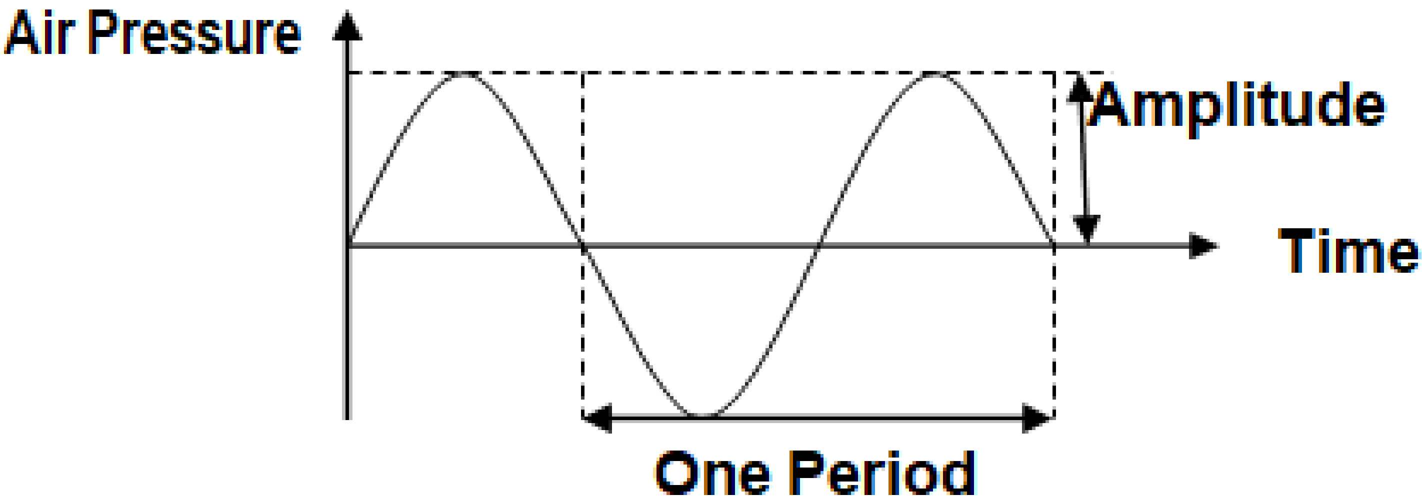


Figure: Oscillation of an air pressure wave

2.1. Basic Sound Concept: Representation and Format…

Frequency

- The frequency of a sound is the reciprocal value of the period. ***It represents the number of periods in a second***. It is measured in hertz (Hz) or cycles per second (cps).

1 KHz = 1000 Hz

Some of the frequency ranges are:

- Infra sound – 0 - 20 Hz
- Human audible sound – 20 Hz - 20KHz
- Ultra sound – 20KHz - 1GHz
- Hyper sound – 1GHz - 10THz

Human audible sound is also called audio or acoustic signals (waves). Speech is an acoustic signal produced by the humans.

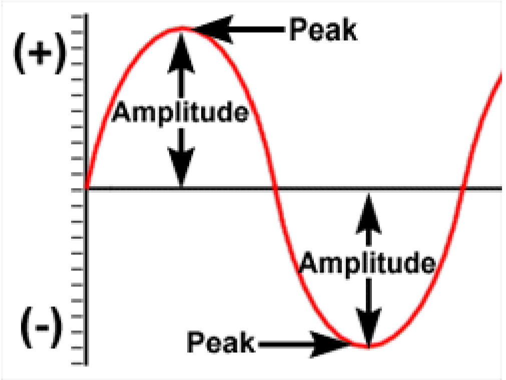
2.1. Basic Sound Concept: Representation and Format…

Amplitude

The amplitude of the sound is the measure of the displacement of the air pressure wave from its mean or quiescent state.

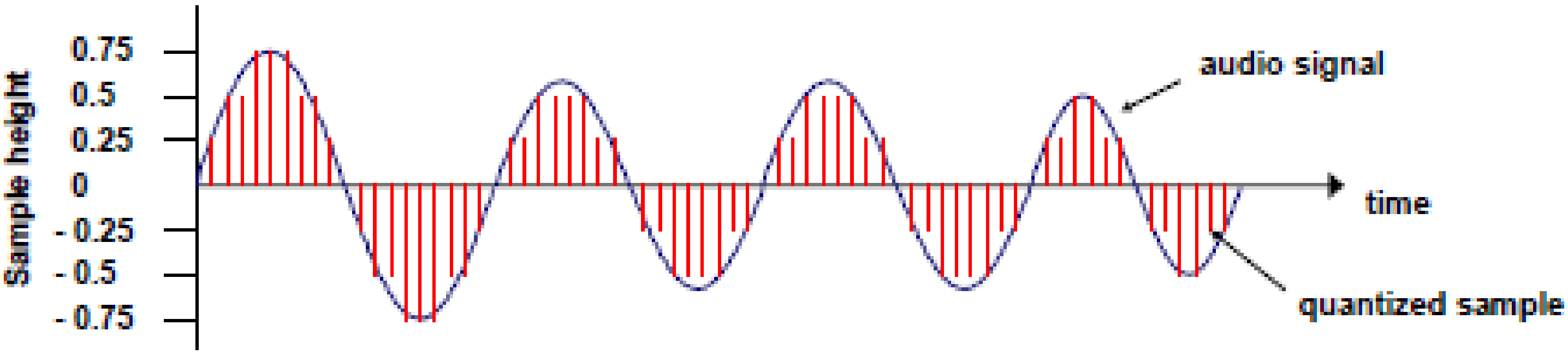
i.e. Distance moved by a point on a vibrating body or wave measured from its equilibrium position.

A sound has an amplitude property subjectively heard as loudness



2.1. Basic Sound Concept: Representation and Format…

Computer representation of sound



- A transducer converts pressure to voltage levels
- The analog signal is converted into a digital stream by discrete sampling:
- The analogous signal is sampled in regular time intervals, i.e. the amplitude of the wave is measured
- Discretization both in time and amplitude (quantization) to get representative values in a limited range (e.g. quantization with 8 bit: 256 possible values)
- Result: series of values:

0.25	0.5	0.5	0.75	0.75	0.75	0.5	0.5	0.25	0	- 0.25	- 0.5	…

2.1. Basic Sound Concept: Representation and Format…

Computer representation of sound

The smooth, continuous curve of a wave form is not directly represented in a computer. ***A computer measures the amplitude of the waveform at regular time intervals to produce a series of numbers called samples.***

Audio signals are converted into digital samples through Analog-to-Digital Converter (ADC). The reverse mechanism is performed by a Digital-to Analog Converter (DAC). E.g. of ADC is AM79C30A digital subscriber controller chip.

2.1. Basic Sound Concept: Representation and Format…

Sampling

Sound wave form the smooth, continuous is not directly represented in the computer.

The computer measures the amplitude of the wave form in the regular time interval to produce the series the numbers. Each of this measurement is called sample. This process is called sampling.

Sampling rate: the rate at which a continuous wave form is sampled is called sampling rate. Like frequency, sampling rate are measured in Hz. For loss less digitization the sampling rate should be at least twice of the maximum frequency response.

2.1. Basic Sound Concept: Representation and Format…

Sampling Rate..

Nyquist Sampling Theorem: “For lossless digitization, the sampling rate should be at least twice the maximum frequency responses”.

E.g. CD standard sampling rate of 44100Hz means that the waveform is sampled 44100 times per second. A sampling rate of 44100 Hz can only represent frequencies up to 22050 Hz.

2.1. Basic Sound Concept: Representation and Format…

Sample Height

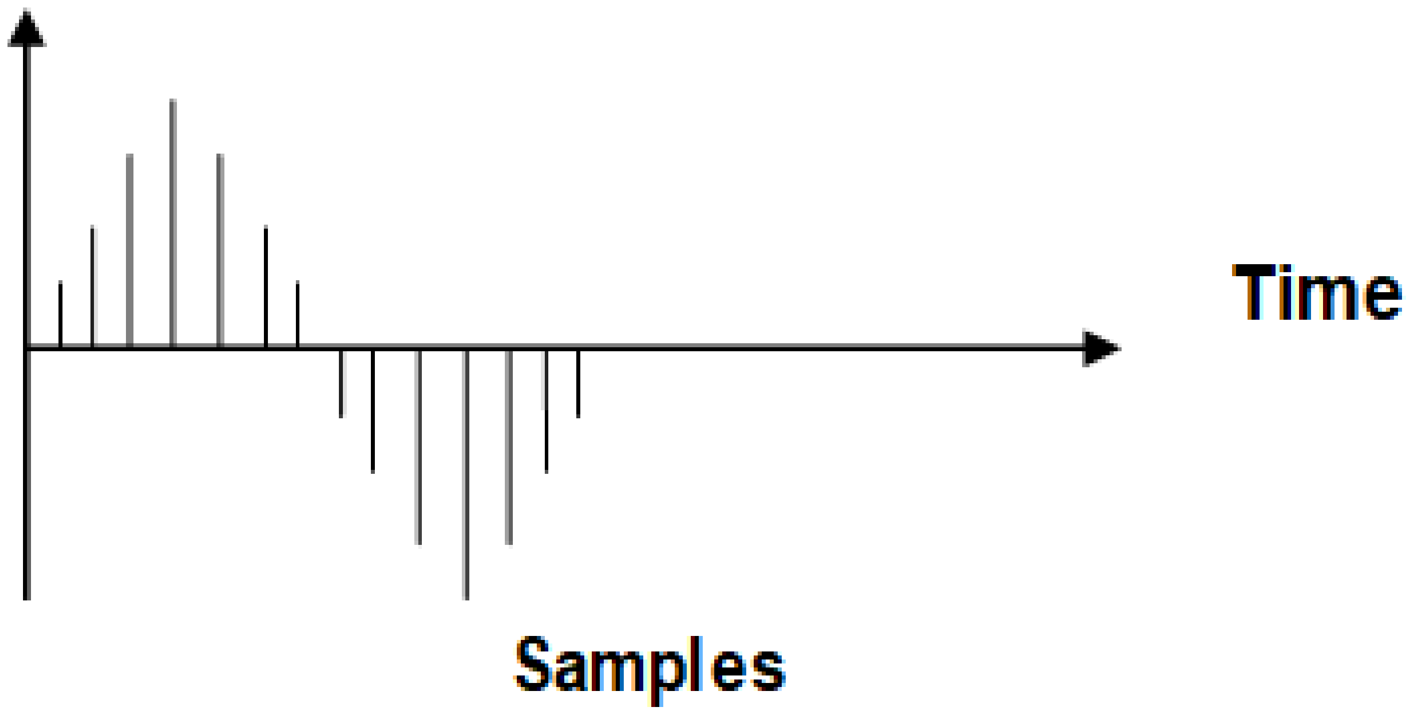


Figure: Sampled Waveform

2.1. Basic Sound Concept: Representation and Format…

Quantization:

The value of sample is discrete. Resolution/Quantization of a sample value depends on the no. of bits used in measuring the height of a waveform.

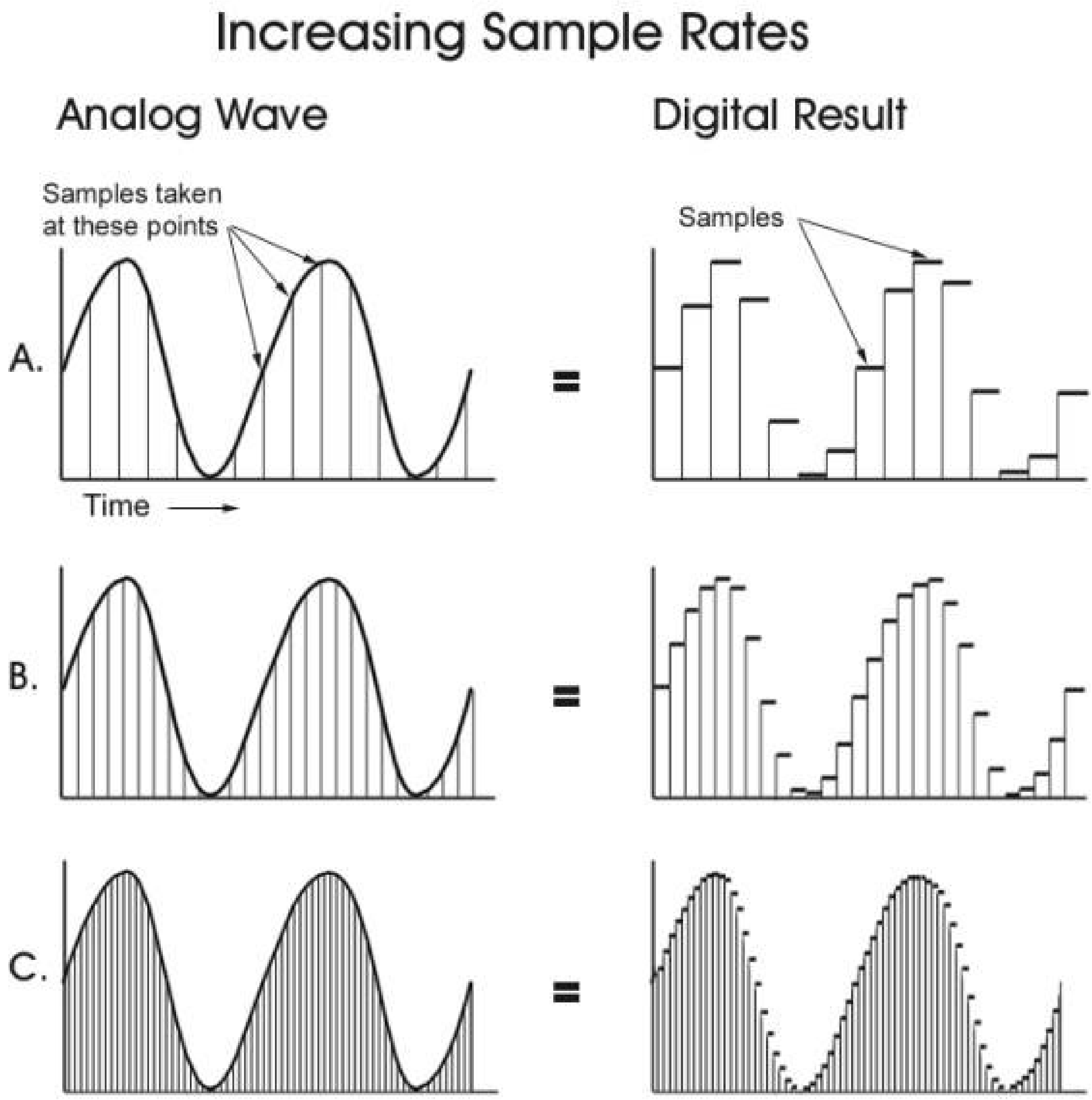
For E.g. An 8-bit quantization yields 256 values. 16-bit CD quality quantization results over 65536 values.

2.1. Basic Sound Concept: Representation and Format…

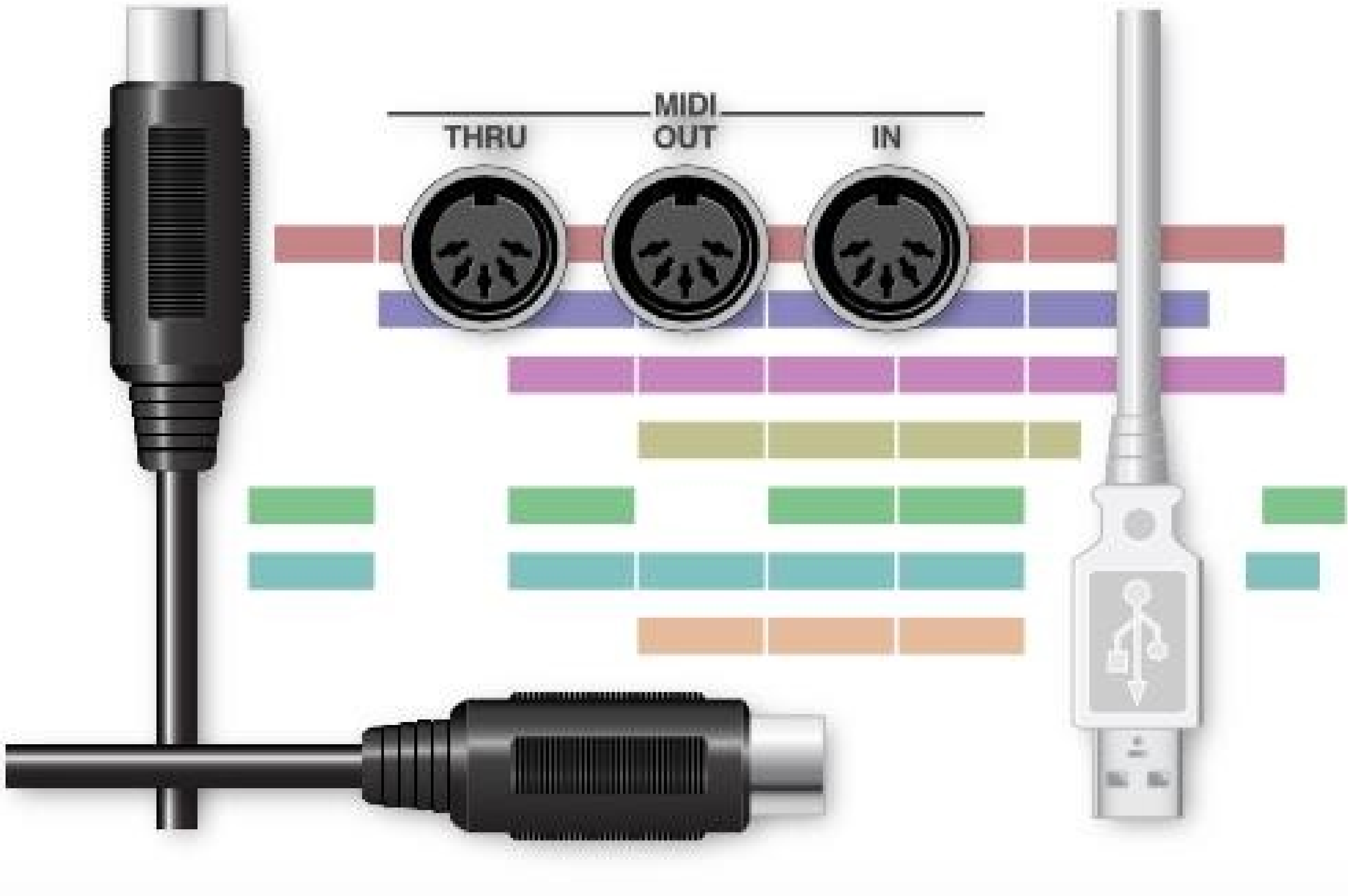
- **Quantization:**

Figure: Three-bit Quantization

Lower the quantization; lower the quality of the sound. For 3-bit quantization, values are 8. i.e. 0.75, 0.5, 0.25, 0, -0.25, -0.5, -0.75 & 1


Increasing Sample Rates
Analog Wave
Digital Result
Samples taken
at these points
Samples
A.
B.
C.
Time
=
=
=

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software



2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

- **Music:**

The relationship between music and computer has become more and more important, especially considering the development of **MIDI (Musical Instrument and Digital Interface)** and its important contribution in the music industry.

MIDI interface between musical instrument and computer.

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

- **MIDI Basic Concept:**

MIDI is the standard that manufacturer of electronic musical instrument have agreed upon.

It is set of specifications used for building the instrument so that the instrument of one manufacturer can without difficulty communicate musical information between one another.

So that **the instrument of different manufacturers can communicate without difficulty**. MIDI is defined in 1983.

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

MIDI Interface Components:

MIDI interface has two different components:

i. Hardware:

ii. A data format:

i. Hardware:

Hardware connects the equipment. It specifies the physical connection between musical instruments.

A MIDI Port is built in to the instrument and it specifies MIDI cable and deals with electronic signals that are sent over the cable.

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

ii. A data format:

Data format encodes the information travelling through the hardware. MIDI doesn't transmit an audio signal.

For this propose MIDI data format is used. The MIDI encoding includes, besides the instrument specification, the notion of the beginning and ending node, basic frequency and sound volume.

MIDI data format is digital. The data are grouped into MIDI message. Each MIDI message can communicate one musical event between instruments.

MIDI: Data Format

- Information traveling through the hardware is encoded in MIDI data format.
- The encoding includes *note* information like beginning of note, frequency and sound volume; upto 128 notes
- The MIDI data format is digital
- The data are grouped into MIDI messages
- Each MIDI message communicates one *musical event* between machines. An event might be pressing keys, moving slider controls,
- 10 mins of music encoded in MIDI data format is about 200 Kbytes of data. (compare against CD-audio!)

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

MIDI Reception Mode:

- Mode 1: Omni on / poly
- Mode 2: Omni on / mono
- Mode 3: Omni off / poly
- Mode 4: Omni off / mono

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

Omni On/Off:

If Omni if turned on, the MIDI device monitors all the MIDI channels and responds to all channels messages. If it is turned off, the MIDI device responds only to channel messages sent on the channels the device is set to receive.

Omni Poly/Mono:

Of Poly is set, the device can play several notes at a time. If the mono is set, the device plays notes like monophonic synthesizer-one note at a time.

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

The first half of the mode name specifies how MIDI device monitors the incoming MIDI channels. If Omni is turned on, the MIDI device monitors all the MIDI channels and responds to all channel messages, no matter which channels they are transmitted on. If the Omni is turned off, the MIDI device respond only to channel message, the device is set to receive.

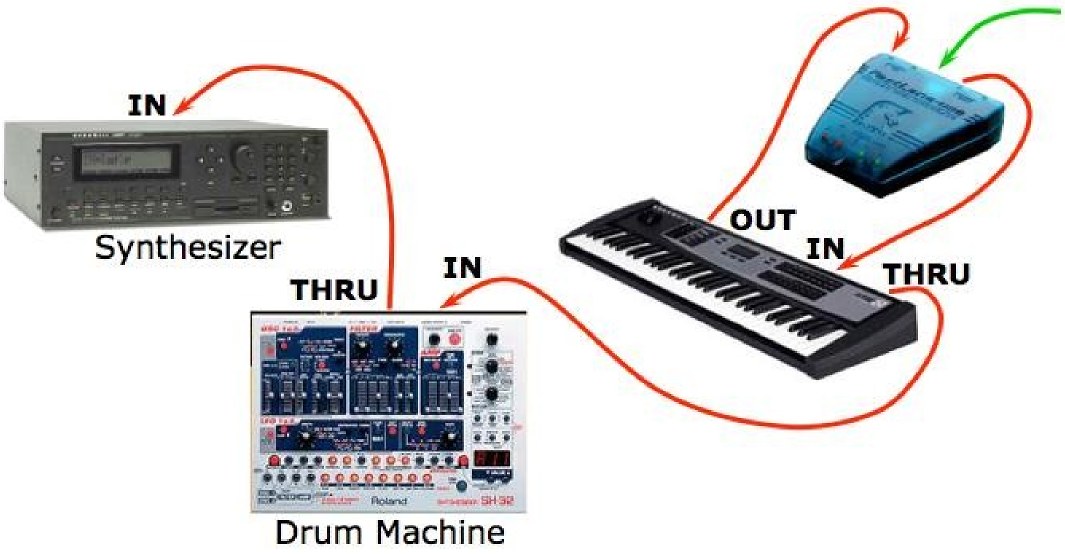
The second half of the mode name tells the MIDI device how to play nodes coming in over the MIDI cable. If the option Poly is set the device can play several nodes at a time. If the node is set to mono the device play only one node at a time like a monophonic synthesizer.

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

MIDI Devices

MIDI synthesizer device is the heart of MIDI system. Most **synthesizers** have following components.

1. ***Sound generators***
2. ***Microprocessor***
3. ***Keyboard***
4. ***Control panel***
5. ***Auxiliary controllers***
6. ***Memory***
7. ***Drum machine***
8. ***Master keyboard***
9. ***Guitar Synthesizer***



2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

- ***Sound generators**:*

  It synthesizes the sound. It produces an audio signal that becomes sound when fed into a loud speaker. It can change quality of sound by varying the voltage oscillation of the audio. Sound generation is done in 2-ways:

  - Storing acoustic signals as MIDI data in advance
  - Creating acoustic signals synthetically

- ***Microprocessor**:*

  Microprocessor communicates with the keyboard to know which notes the musician is playing. Microprocessor communicates with the control panel to know what commands the musician wants to sent to the microprocessor. The microprocessor then specifies note and sound commands to the sound generators (i.e. microprocessor sends and receives the MIDI message).

- ***Keyboard**:*

  It affords the musician‟s direct control of the synthesizer. Pressing keys means signaling microprocessor what notes to play and how long to play them. Keyboard should have at least 5 octaves and 61 keys.

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

- ***Control panel**:*

  Controls those function that are not directly concerned with notes and duration. Control panel includes a slider, a button and a menu.

- ***Auxiliary controllers**:*

  Gives more control over the notes played on keyboard. Pitch bend and modulation are the 2 common variables on the synthesizer

- ***Memory**:*

  Stores patches for the sound generation and settings on the control panel.

- ***Drum machine**:*

  Specialize in percussion sounds and rhythms.

- ***Master keyboard**:*

  Increases the quality of the synthesizer

- ***Guitar Synthesizer**:*

  Drum pad controllers, Guitar controllers and many more

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

Sequencer:

It is an electronic device in cooperating with both hardware and software, which is used as storage server for generated MIDI data. A sequencer may be computer. Sequencer transforms the note into MIDI message.

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

MIDI Message:

MIDI messages transmit information between MIDI devices and determine what kinds of musical events can be passed from device to device.

Formats of MIDI messages

- ***Status byte***: First byte of any MIDI message. It describes the kind of message
- ***Data byte***: The following bytes.

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

There are two types of MIDI messages

1. **Channel Message** (: *It go only to specified devices*)
   1. Channel Voice Message
   2. Channel mode Message
2. **System Message** (: *It go to all devices in a MIDI system*)
   1. System Real-time message
   2. System Common Message
   3. System exclusive Message

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

1. Channel Message: It go only to specified devices. There are two types of channel message.

i. **Channel Voice Message**: (*send actual performance between devices*)

Channel Voice message send actual performance data between MIDI devices describing keyboard action, controller action and control panel changes. E.g. note on, Note off, channel pressure, control change etc.

ii. **Channel Mode Message**: (*determine how to play notes*)

Channel Mode message determine the way that a receiving MIDI device respond to channel voice message. It deals with how to play notes coming in over MIDI cables. Channel mode message includes Omni On, Omni Off, note off, note on, etc. E.g. local control, All note off, Omni mode off etc.

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

2. System message: System message go to all devices in a MIDI system because no channel number are specified. There are three types of system message.

i. **System Real-time Message: (***avoid delay time***)**

These are very short and simple, consisting of only one byte. They carry extra data with them. ***These messages synchronize the timing MIDI devices in performance. To avoid delay these message are sent in the middle of other message if necessary***. E.g. System reset, Timing clock i.e. MIDI clock etc.

ii. **System Common Message: (***prepare sequence***)**

System common message are commands that ***prepares sequencers and synthesizer to play song***. E.g. Song selected, find the common starting place in the song.

iii. **System Exclusive message:** (*create customize message*)

System exclusive message allow MIDI manufacturers to ***create customized MIDI message to send between the MIDI devices.***

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

MIDI Software

The software application generally falls into four major categories.

1. Music recording and performance application
2. Musical notation and printing application
3. Synthesizer Patch editor and library patch
4. Music education application

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

1. ***Music recording and performance applications:*** Provides function as recording of MIDI messages. Editing and playing the messages in performance.
2. ***Musical notations and printing applications:*** Allows writing music using traditional musical notation. User can play and print music on paper for live performance or publication.
3. ***Synthesizer path editor and librarians:*** Allows information storage of different synthesizer patches in the computer's memory and disk drives. Editing of patches in computer.
4. ***Music education applications:*** Teaches different aspects of music using the computer monitor, keyboard and other controllers of attached MIDI instruments.

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

MIDI and SMPTE Timing Standard

MIDI reproduces traditional note length using MIDI clock, which are represented through timing clock message. Using MIDI clock a receiver can synchronize with the clock cycle of sender. To keep the standard timing reference the MIDI specification state **24 MIDI clock = 1 quarter note**.

As an alternative the SMPTE timing standard (***Society Of Motion Picture and Television Engineer***) can be used. The SMPTE timing standard was originally developed by NASA as a way to make incoming data from different tracking stations so that the receiving computer could fill what time each piece of data was created. SMPTE format consists of hour: minutes: second: frames: bits

-30 frames per second, SMPTE uses a 24 hour clock from 0 to 23 before recycling.

2.2.Basic Music (MIDI) Concept: Devices, Messages, Standards and Software

Processing chain of interactive computer music systems

- ***Sensing stage:*** Data is collected from controllers reading the gesture information from human performers on stage.
- ***Processing stage**:* Computer reads and interprets information coming from the sensors and prepares data for the response stage.
- ***Response stage:*** Computer ad some collection of sound producing devices share in realizing a musical output

Some of the interactive music systems:

- Max (OOP language)
- Cypher (It has a listener and a player)
- NeXT Computer (It has a music kit)
- M & Jam Factory (It has graphics control panel)

2.3. Speech : Generation, Analysis and Transmission

Speech

Speech can be „perceived", „understood", and „generated" by humans and by machines too. Human speech signal comprises a subjective lowest spectral component known as the pitch. Pitch is not proportional to the frequency. Human ear is sensitive in range from 600Hz-6000Hz.

So, a human adjust himself/ herself very efficiently to different speakers and their speech habit. The human brain can recognize the very fine line between speech and noise.

2.3. Speech : Generation, Analysis and Transmission

Properties of Speech Signals:

1. **Voice speech signals** show during certain time intervals almost periodic behavior. Therefore we can consider these signals as ***quasi-stationary*** signals for around 30ms
2. **The spectrum of the audio signals** shows characteristics maxima, which are mostly 3-5 frequency bands. These maxima called format.

2.3. Speech : Generation, Analysis and Transmission

Speech Generation:

Mid 19th century, Helmholtz built a mechanical vocal tract coupling together several mechanical resonators with which sound could be generated. In 1940, Dudley produced the 1st speech synthesizer through imitation of mechanical vibration using electrical oscillation.

Requirement for speech generation is real time signal generation.

1. Speech output system could transfer text into speech automatically without any lengthy processing.
2. Generated speech must be understandable and must sound natural.

2.3. Speech : Generation, Analysis and Transmission

Basic Notions:

- A ***phone*** is the smallest speech unit, such as the m of mat and the b of bat in English that distinguishes one utterance or word from another in a given language.
- ***Allophones*** mark the variants of a phone. For e.g. the aspirated p of pit and the un-aspirated p of spit are allophones of the English phoneme p.
- The ***morph*** marks the smallest speech unit which carries a meaning itself. Therefore, consider is a morph, but reconsideration is not.
- The ***voiced sound*** is generated through the vocal cords, m, v, l are the examples of voiced sounds.
- During the generation of an ***unvoiced sound*** the vocal cords are opened. F and s are unvoiced sounds.

2.3. Speech : Generation, Analysis and Transmission

Types of Speech generation:

The important requirement of speech generation is the generation of the real time signals. The easiest method for speech generation is to use pre-coded speech and play it back in the timely fashion.

2.3. Speech : Generation, Analysis and Transmission

Time-dependent sound concatenation:

CRUM

Individual speech units are composed like building blocks, where the composition can occur at different levels. Individual phones are understood as speech units. Example, **CRUM** phones word phone are shown individually as -

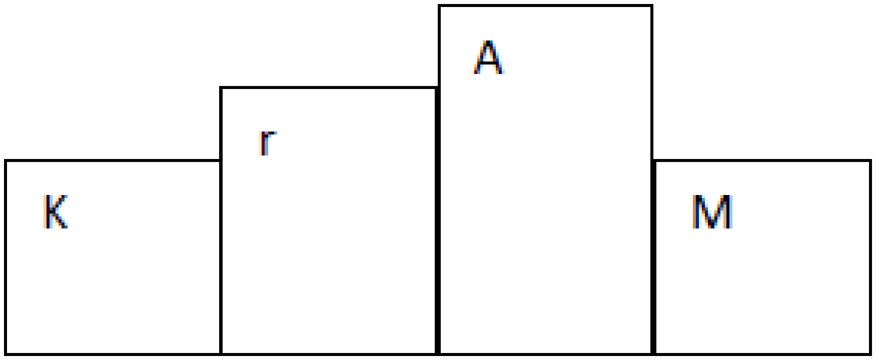


Fig: Phone sound concatenation

2.3. Speech : Generation, Analysis and Transmission

Time-dependent sound concatenation:

CRUM

It is possible with just a few phones to create from an unlimited vocabulary. The phone sound concatenation shows the problem during the transition between individual phones. This problem is called co articulation which is ***mutual sound effect***. To solve these problems, Di-phone sound concatenation is used. Two phones can constitute a Di-phone. In the above figure, Di-phone of word CRUM is shown.

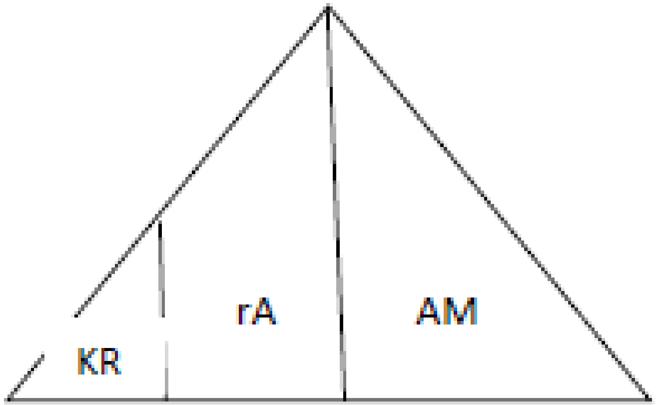


Fig: Di-phone concatenation

2.3. Speech : Generation, Analysis and Transmission

Time-dependent sound concatenation:

CRUM

The transition problem is not solved sufficiently on this level. To make the transition problem easier, syllabus can be created. The speech is generated through set of syllabus. The above given figure word sound concatenation and syllabus sound shows the syllabus sound of word CRUM. The best pronunciation of word is achieved is storage of whole word. This leads towards synthesize the speech sequence.

2.3. Speech : Generation, Analysis and Transmission

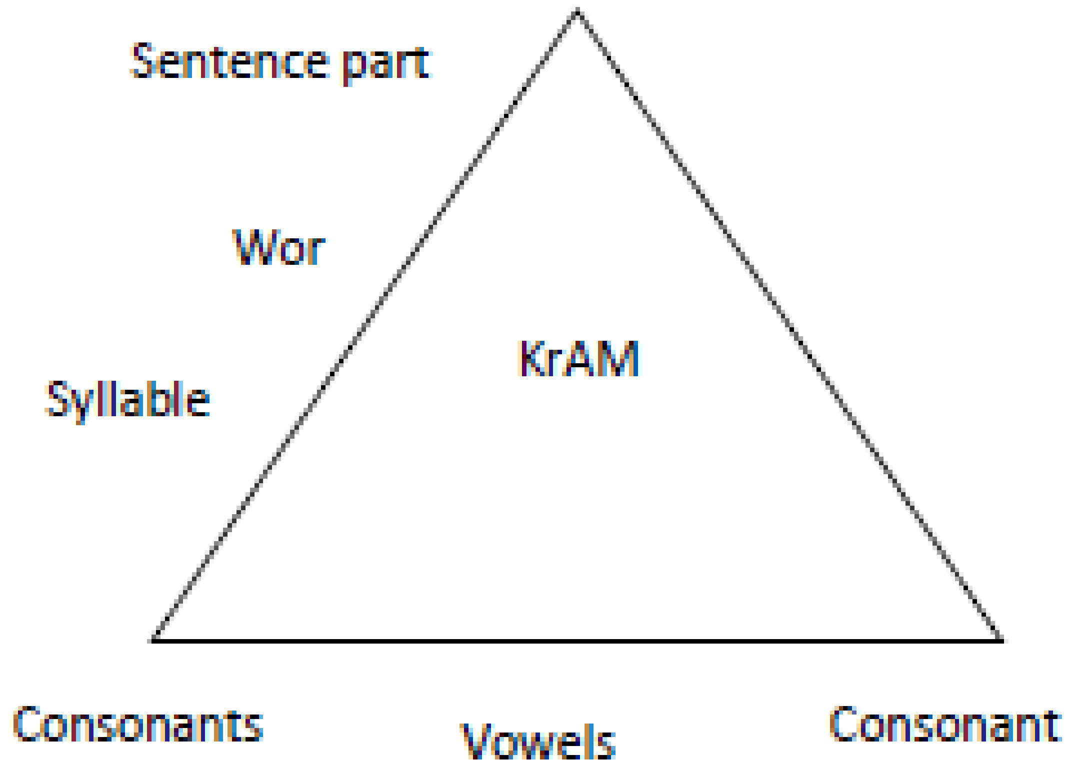


Fig: Word sound concatenation

2.3. Speech : Generation, Analysis and Transmission

- **co articulation**

The transition between individual sound units create an essential problem called co articulation which is the mutual sound influence throughout the several sound.

2.3. Speech : Generation, Analysis and Transmission

Frequency dependent sound concatenation:

Speech generation/output can also be based on a frequency dependent sound concatenation.

This can be done through a ***formant synthesis***. Formants are frequency maxima in the spectrum of the speech signal. Formants synthesis simulates the vocal tract through a filter.

A pulse signal with a frequency, corresponding to the fundamental speech frequency, is chosen as a simulation for voiced sounds. Unvoiced sounds are created through a noise generator.

The method used for the sound synthesis in order to simulate human speech is called the linear predictive coding (LPC) method. Using speech synthesis, an existent text can be transformed into an acoustic signal.

2.3. Speech : Generation, Analysis and Transmission

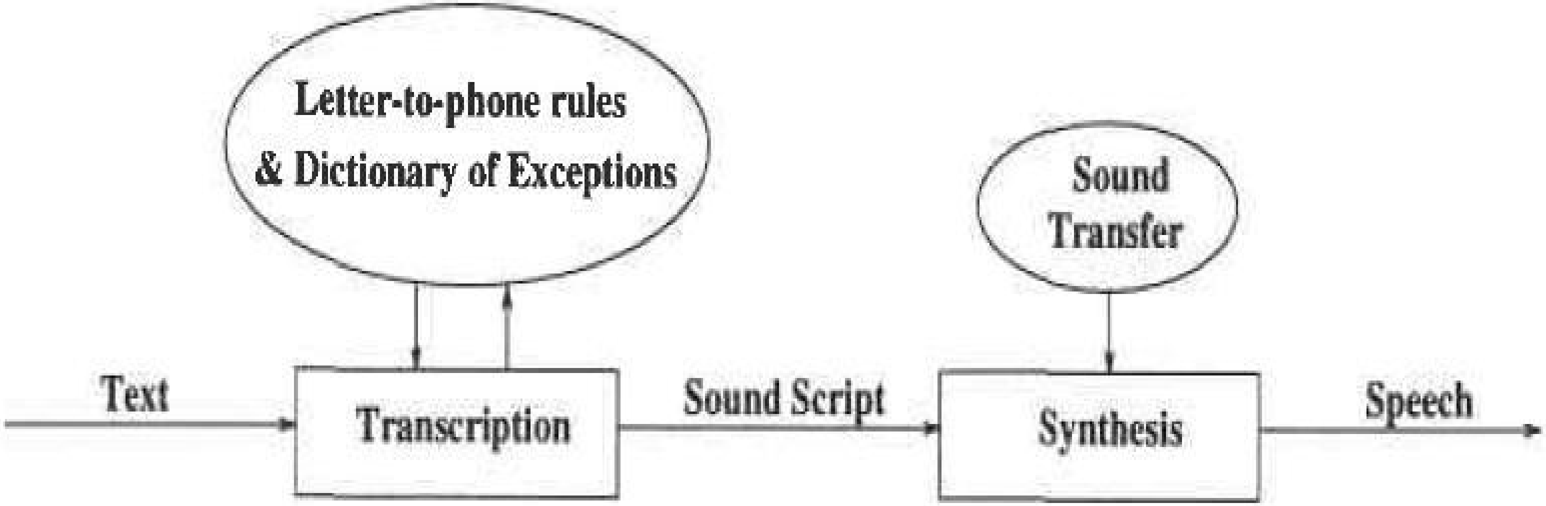


Fig: Components of a speech synthesis system with time-dependent sound concatenation.

In the first step the transcription is performed where text is translated into sound script. Most transcription methods work with later to phone rules. A dictionary of exception stored in the library.

In the second step the sound script is translated into a speech signals. Besides the problem of co-articulation ambiguous pronunciation most be considered.

2.3. Speech : Generation, Analysis and Transmission

Step 1:

- Performs transcription
- Text is translated into sound script
- This process is done using letter-to-phone rules and dictionary of exceptions
- User recognizes the formula deficiency in the transcription and improves the pronunciation manual

Step 2:

- Sound script is translated into a speech signal.
- Time or frequency dependent concatenation can follow

2.3. Speech : Generation, Analysis and Transmission

Speech Analysis

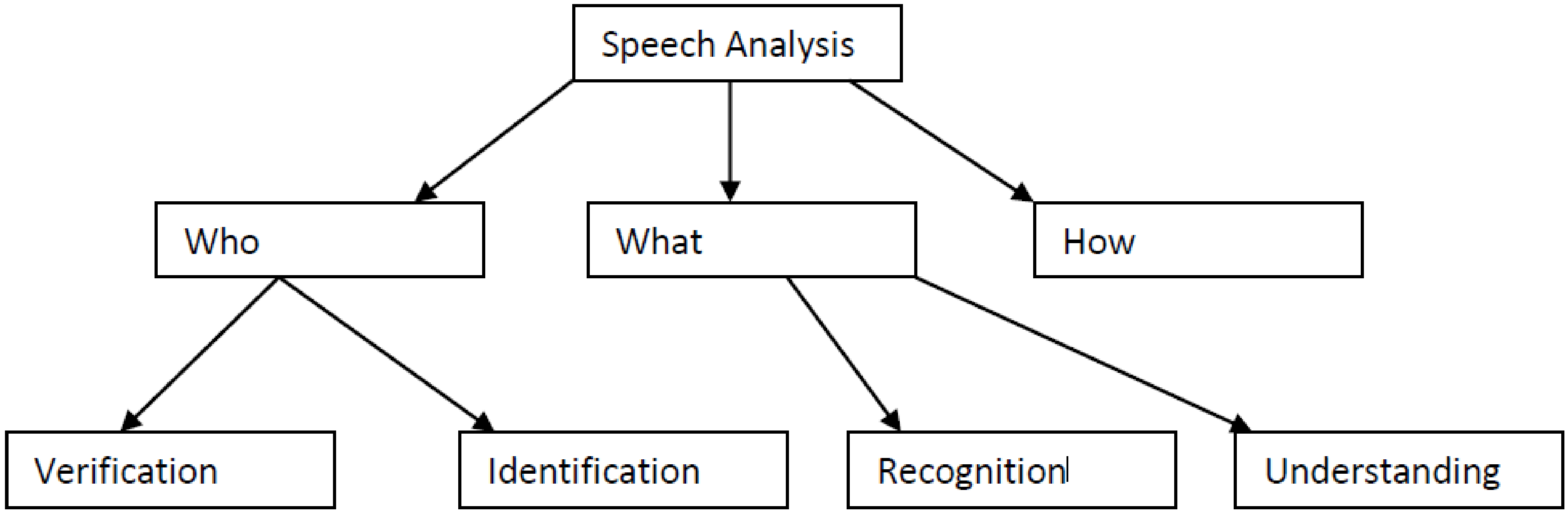


Fig. Speech analysis deals with research area shown in the above figure.

2.3. Speech : Generation, Analysis and Transmission

Speech Analysis

- Human speeches have certain characteristics determined by speaker. So speech analysis can server to analyze who is speaking. i.e. to recognize a speaker, for his identification and verification. The computer identifies and verifies the speaker using an acoustic finger print is digitally stored. It is digitally stored speech probe (certain statement of the speaker).
- Another main part of speech analysis is to analyze what has been said. To recognize and understand the speech itself.
- Another area of speech analysis tries to research speech pattern with respect to how the statement was said. E.g. a speaker sentence sounds differently if a person is angry or happy.

2.3. Speech : Generation, Analysis and Transmission

Speech Recognition System and Understanding:

The system which provides the recognition and understanding of speech signal applies this principle several times as follows:

1. In the first step, the principle is applied to a sound pattern and/or word model. An acoustic and **phonetic analysis is performed.**
2. In the second step, certain speech units go through **syntactical analysis**. In this step the errors in the previous step can be recognized.
3. The third step deals with **semantics(the study of meanings**) of the previously

2.3. Speech : Generation, Analysis and Transmission

Speech Recognition System and Understanding:

There are still many problems into speech recognition and understanding research.

1. Specific problem is presented by Room Acoustic with existed environment noise.
2. Word boundaries must be determined. Very often neighboring words follows into one another.
3. For the comparison of the speech elements to the existing pattern, time normalization is necessary. The same word can be spoken quickly or slowly.

2.3. Speech : Generation, Analysis and Transmission

Speech Transmission

The area of speech transmission deals with efficient coding of the speech signal to allow speech/sound transmission at low transmission rates over networks.

The goal is to provide the receiver with the same speech/sound quality as was generated at the sender side. This section includes some principles that are connected to speech generation and recognition'

2.3. Speech : Generation, Analysis and Transmission

<u>Speech Transmission</u>

Signal Form Coding

This kind of coding considers no speech-specific properties and parameters. Here, the goal is to achieve the most efficient coding of the audio signal. The data rate of a PCM-coded stereo-audio signal with CD-quality requirements is:

Rate = 2 * 44100/s * 16 bit / 8 bit/byte

= 176400b bytes/s = 1411200 bits / s

Telephone quality, in comparison to CD-quality, needs only 64 Kbit/s. Using Difference Pulse Code Modulation (DPCM), the data rate can be lowered to 56 Kbits/s without loss of quality. Adaptive Pulse Code Modulation (ADPCM) allows a further rate reduction to 32 Kbits/s.

2.3. Speech : Generation, Analysis and Transmission

Speech Transmission

Source Coding

Parameterized systems work with source coding algorithms. Here, the specific speech characteristics are used for data rate reduction. Channel vo-coder is an example of such a parameterized system.

The channel vo-coder is an extension of a sub-channel coding. The signal is divided into a set of frequency channels during speech analysis because only certain frequency maxima are relevant to speech.

A **vocoder** is a category of voice codec that analyzes and synthesizes the human voice signal for audio data compression, multiplexing, voice encryption, voice transformation, etc.

vocoder

2.3. Speech : Generation, Analysis and Transmission

Speech Transmission

Source Coding

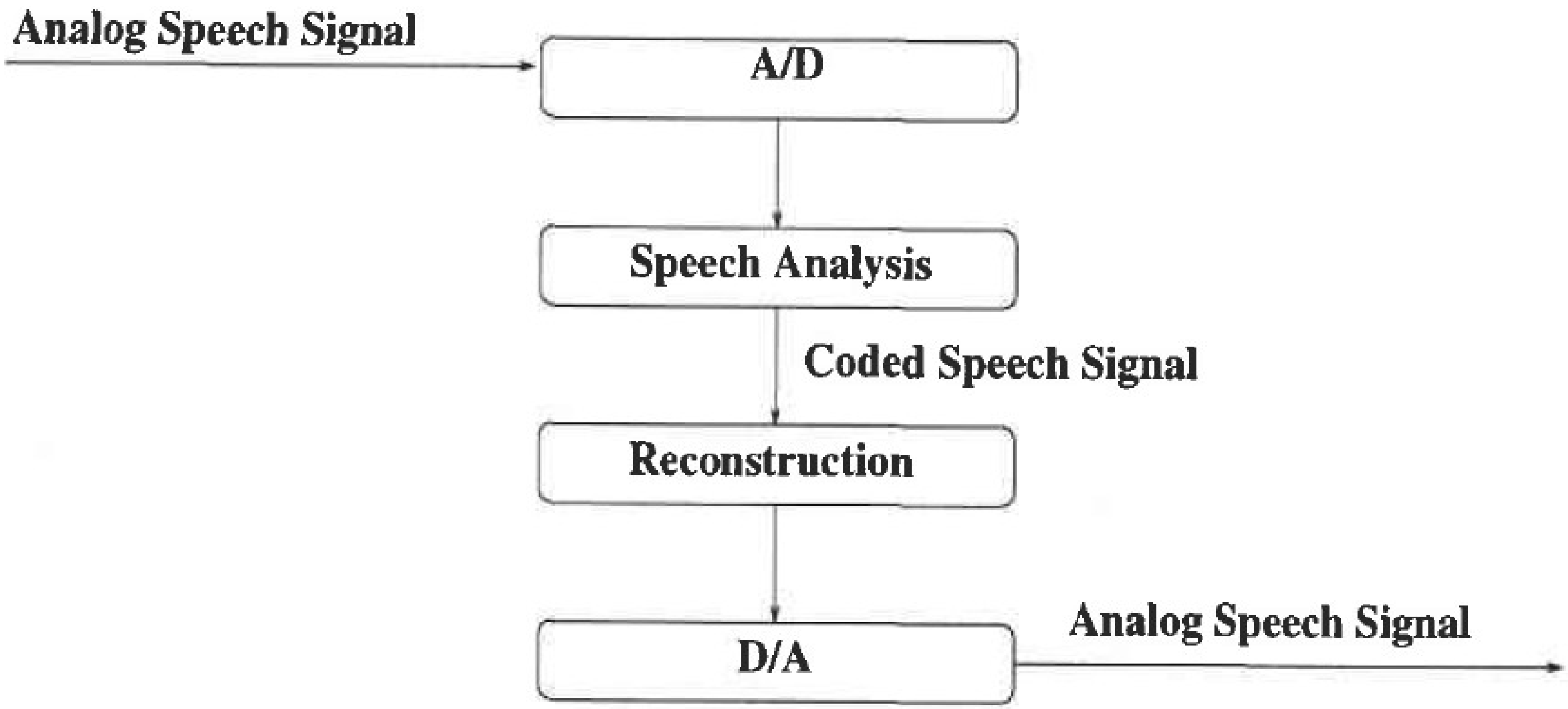


Fig: Source coding in parameterized systems: components of a speech transmission system.

2.3. Speech : Generation, Analysis and Transmission

Speech Transmission

Recognition / Synthesis (Mixture) Methods

There have been attempts to reduce the transmission rate using pure recognition/ synthesis methods. Speech analysis (recognition) follows on the sender side of a speech transmission system and speech synthesis (generation) follows

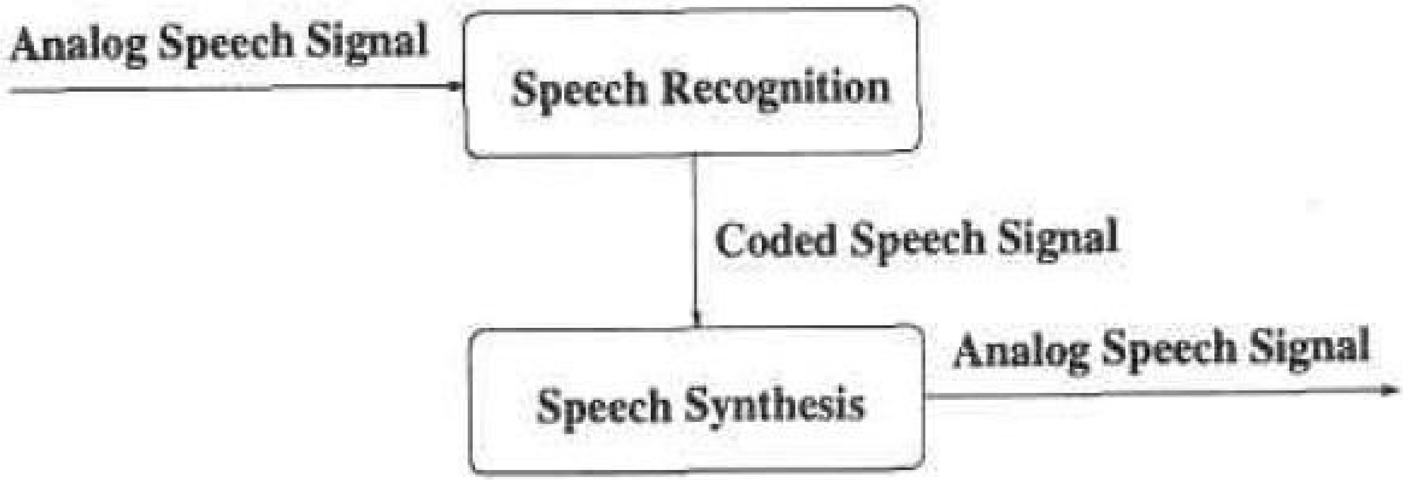


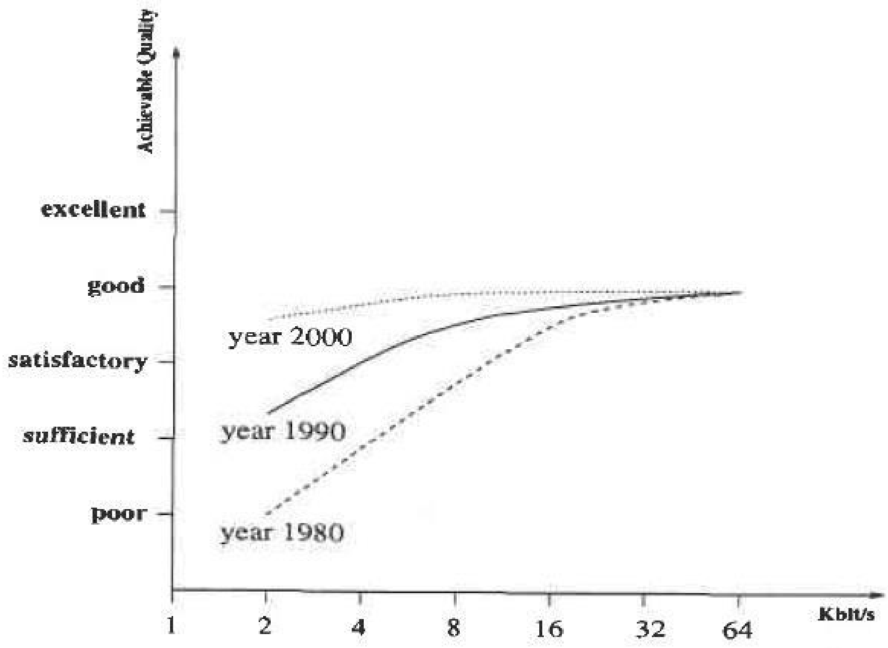
Fig: Recognition/synthesis systems: components of a speech transmission system.

2.3. Speech : Generation, Analysis and Transmission

Speech Transmission

Achieved Quality

The essential question regarding speech and audio transmission with respect to multimedia systems is how to achieve the minimal data rate for a given quality.



References:

- Multimedia: Computing, Communications and Applications", Ralf Steinmetz and Klara Nahrstedt, Pearson Education Asia
- "Multimedia Communications, Applications, Networks, protocols ad Standards", Fred Halsall, Pearson Education Asia
- "Multimedia Systems", John F. Koegel Buford, Pearson Education Asia