MPEG Audio Compression

14.1 Psychoacoustics

- The range of human hearing is about 20 Hz to about 20 kHz.
- The frequency range of the voice is typically only from about 500 Hz to 4 kHz.
- The dynamic range, the ratio of the maximum sound amplitude to the quietest sound that humans can hear, is on the order of about 120 dB.

Equal-Loudness Relations

- **Fletcher-Munson Curves**
  - Equal loudness curves that display the relationship between perceived loudness ("Phons", in dB) for a given stimulus sound volume ("Sound Pressure Level", also in dB), as a function of frequency.
- Fig. 14.1 shows the ear's perception of equal loudness:
  - The bottom curve, for example, shows what level of pure tone stimulus is required to produce the perception of a 10 dB sound.
  - All the curves are arranged so that the perceived loudness level gives the same loudness as for that loudness level of a pure tone at 1 kHz.

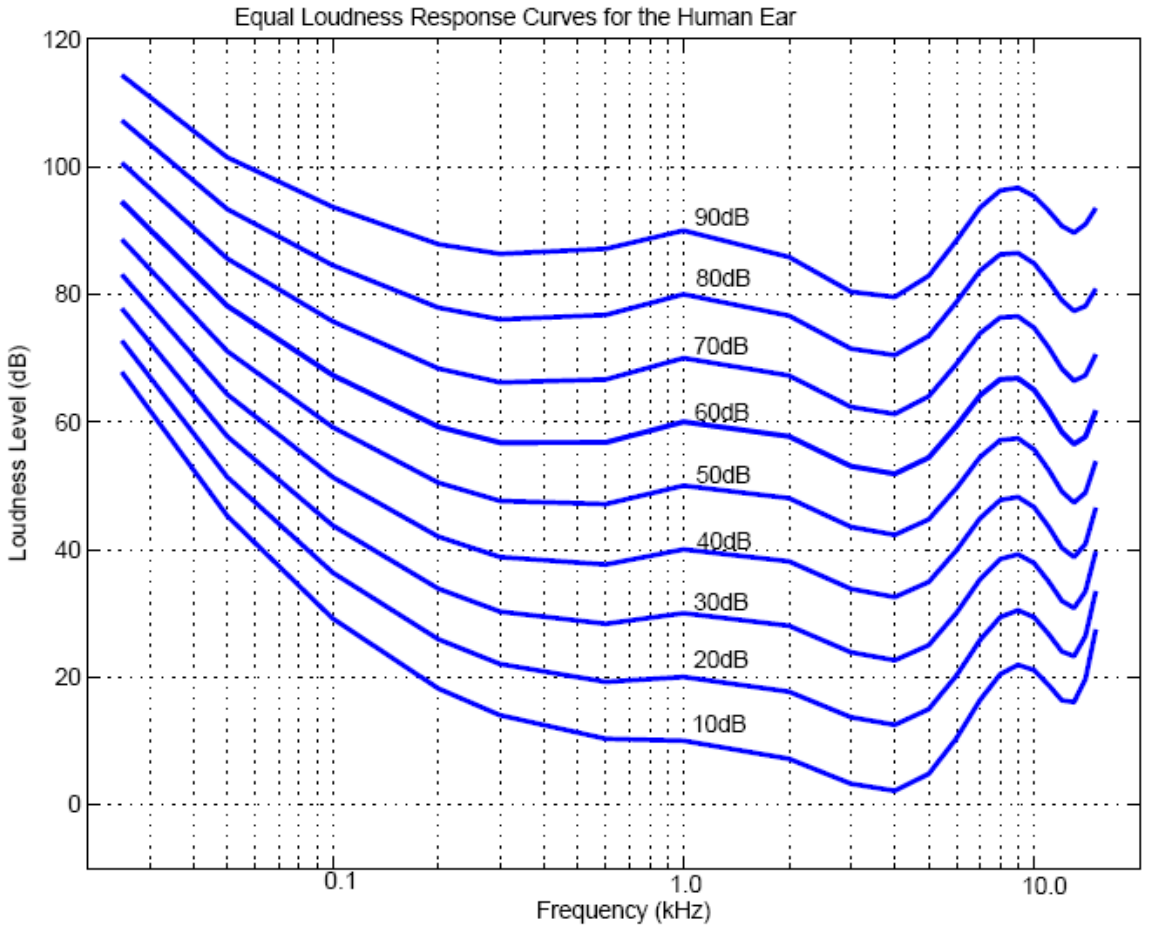


Fig. 14.1: Fletcher-Munson Curves (re-measured by Robinson and Dadson)

Frequency Masking

- Lossy audio data compression methods, such as MPEG/Audio encoding, remove some sounds which are masked anyway.
- The general situation in regard to masking is as follows:

1. A lower tone can effectively mask (make us unable to hear) a higher tone.
2. The reverse is not true – a higher tone does not mask a lower tone well.
3. The greater the power in the masking tone, the wider is its influence – the broader the range of frequencies it can mask.
4. As a consequence, if two tones are widely separated in frequency then little masking occurs.

Threshold of Hearing

- A plot of the threshold of human hearing for a pure tone

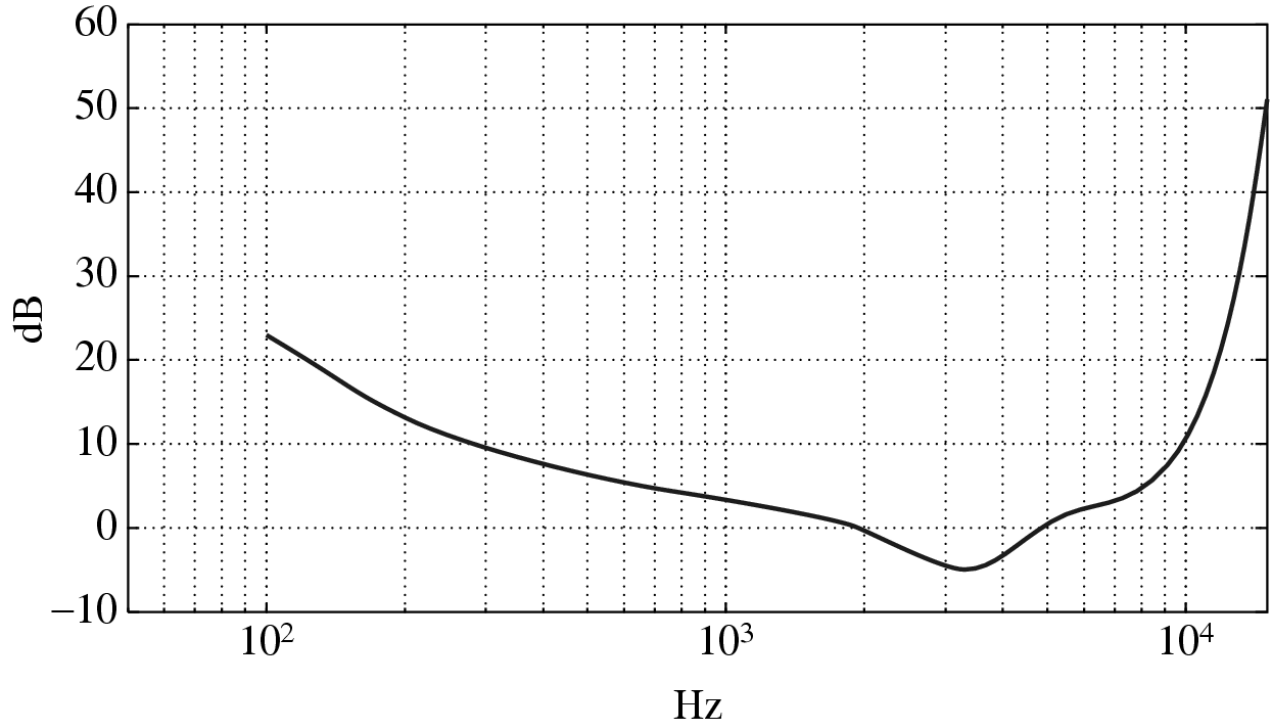


Fig. 14.2: Threshold of human hearing, for pure tones

Threshold of Hearing (Cont'd)

- The threshold of hearing curve: if a sound is above the dB level shown then the sound is audible.
- Turning up a tone so that it equals or surpasses the curve means that we can then distinguish the sound.
- An approximate formula exists for this curve:

$$\text{Threshold}(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4 \quad (14.1)$$

- The threshold units are dB; the frequency for the origin (0,0) in formula (14.1) is 2,000 Hz: $\text{Threshold}(f) = 0$ at f =2 kHz.

Frequency Masking Curves

- Frequency masking is studied by playing a particular pure tone, say 1 kHz, at a loud volume, and determining how this tone affects our ability to hear tones nearby in frequency.
  - One would generate a 1 kHz *masking* tone, at a fixed sound level of 60 dB, and then raise the level of a nearby tone, e.g., 1.1 kHz, until it is just audible.
- The threshold in Fig. 14.3 plots the audible level for a single masking tone (1 kHz).
- Fig. 14.4 shows how the plot changes if other masking tones are used.

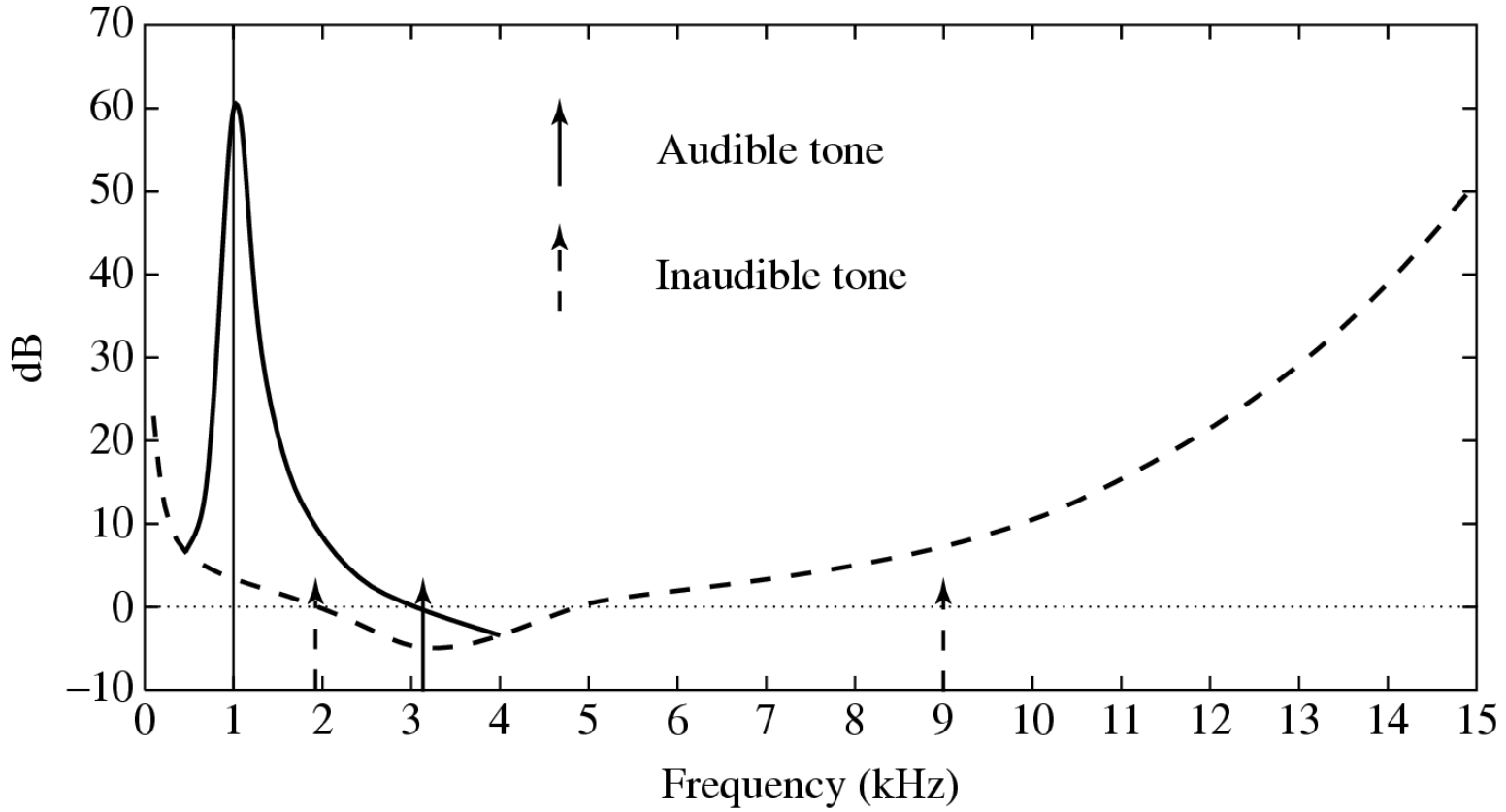


Fig. 14.3: Effect on threshold for 1 kHz masking tone

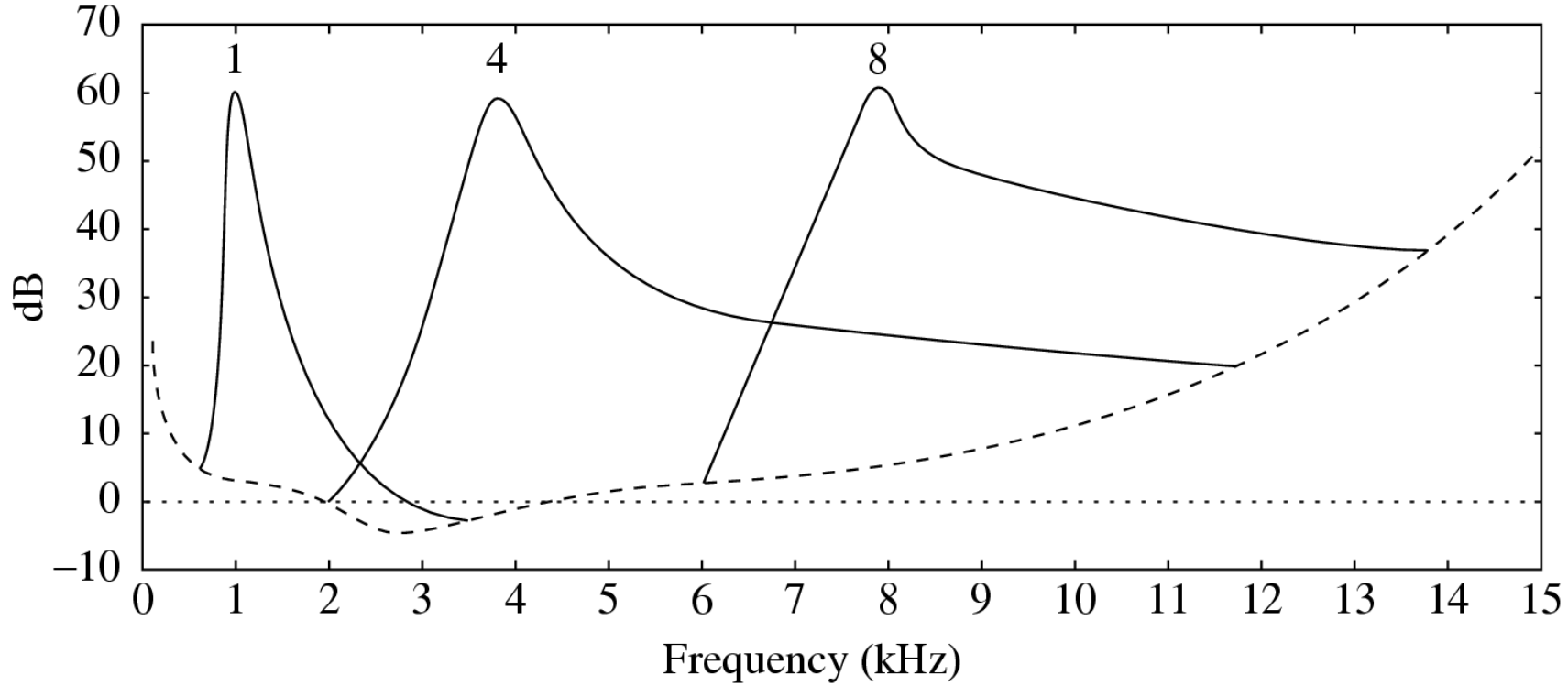


Fig. 14.4: Effect of masking tone at three different frequencies

Critical Bands

- **Critical bandwidth** represents the ear's resolving power for simultaneous tones or partials.
  - At the low-frequency end, a critical band is less than 100 Hz wide, while for high frequencies the width can be greater than 4 kHz.
- Experiments indicate that the critical bandwidth:
  - for masking frequencies < 500Hz: remains approximately constant in width (about 100 Hz)
  - For masking frequencies > 500Hz: increases approximately linearly with frequency.

Table 14.1: 25 Critical Bands and Their Bandwidths

Band #	Lower Bound (Hz)	Center (Hz)	Upper Bound (Hz)	Bandwidth (Hz)
1	-	50	100	-
2	100	150	200	100
3	200	250	300	100
4	300	350	400	100
5	400	450	510	110
6	510	570	630	120
7	630	700	770	140
8	770	840	920	150
9	920	1000	1080	160
10	1080	1170	1270	190
11	1270	1370	1480	210
12	1480	1600	1720	240

Band #	Lower Bound (Hz)	Center (Hz)	Upper Bound (Hz)	Bandwidth (Hz)
13	1720	1850	2000	280
14	2000	2150	2320	320
15	2320	2500	2700	380
16	2700	2900	3150	450
17	3150	3400	3700	550
18	3700	4000	4400	700
19	4400	4800	5300	900
20	5300	5800	6400	1100
21	6400	7000	7700	1300
22	7700	8500	9500	1800
23	9500	10500	12000	2500
24	12000	13500	15500	3500
25	15500	18775	22050	6550

Bark Unit

- **Bark unit** is defined as the width of one critical band, for any masking frequency.
- The idea of the Bark unit: every critical band width is roughly equal in terms of Barks (refer to Fig. 14.5).

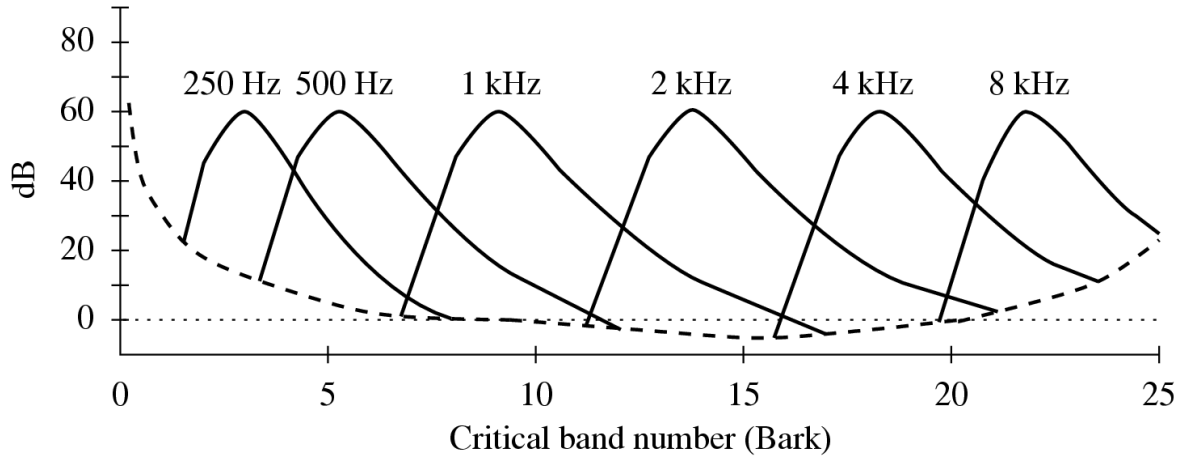


Fig. 14.5: Effect of masking tones, expressed in Bark units

Conversion: Frequency & Critical Band Number

- Conversion expressed in the Bark unit:

$$\text{Critical band number (Bark)} = \begin{cases} f/100, & \text{for } f < 500 \\ 9 + 4\log_2(f/1000), & \text{for } f \geq 500 \end{cases} \quad (14.2)$$

- Another formula used for the Bark scale:

$$b = 13.0\arctan(0.76f) + 3.5\arctan(f^2/56.25) \quad (14.3)$$

where f is in kHz and b is in Barks (the same applies to all below).

- The inverse equation:

$$f = [(\exp(0.219 * b)/352) + 0.1] * b - 0.032 * \exp[-0.15 * (b - 5)2] \quad (14.4)$$

- The critical bandwidth (df) for a given center frequency f can also be approximated by:

$$df = 25 + 75 \times [1 + 1.4(f^2)]^{0.69} \quad (14.5)$$

Temporal Masking

- **Phenomenon**: any loud tone will cause the hearing receptors in the inner ear to become *saturated* and require time to recover
- The following figures show the results of masking experiments:

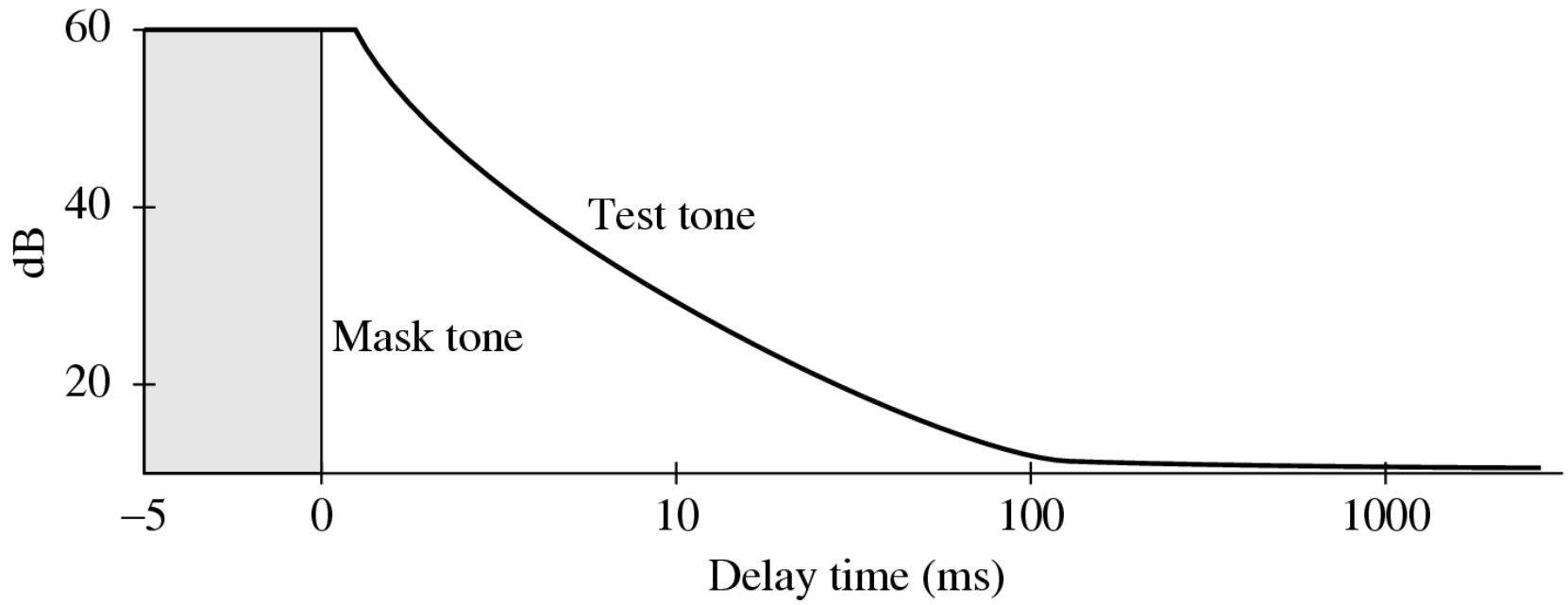


Fig.14.6: The louder the test tone, the shorter the amount of time required before the test tone is audible once the masking tone is removed.

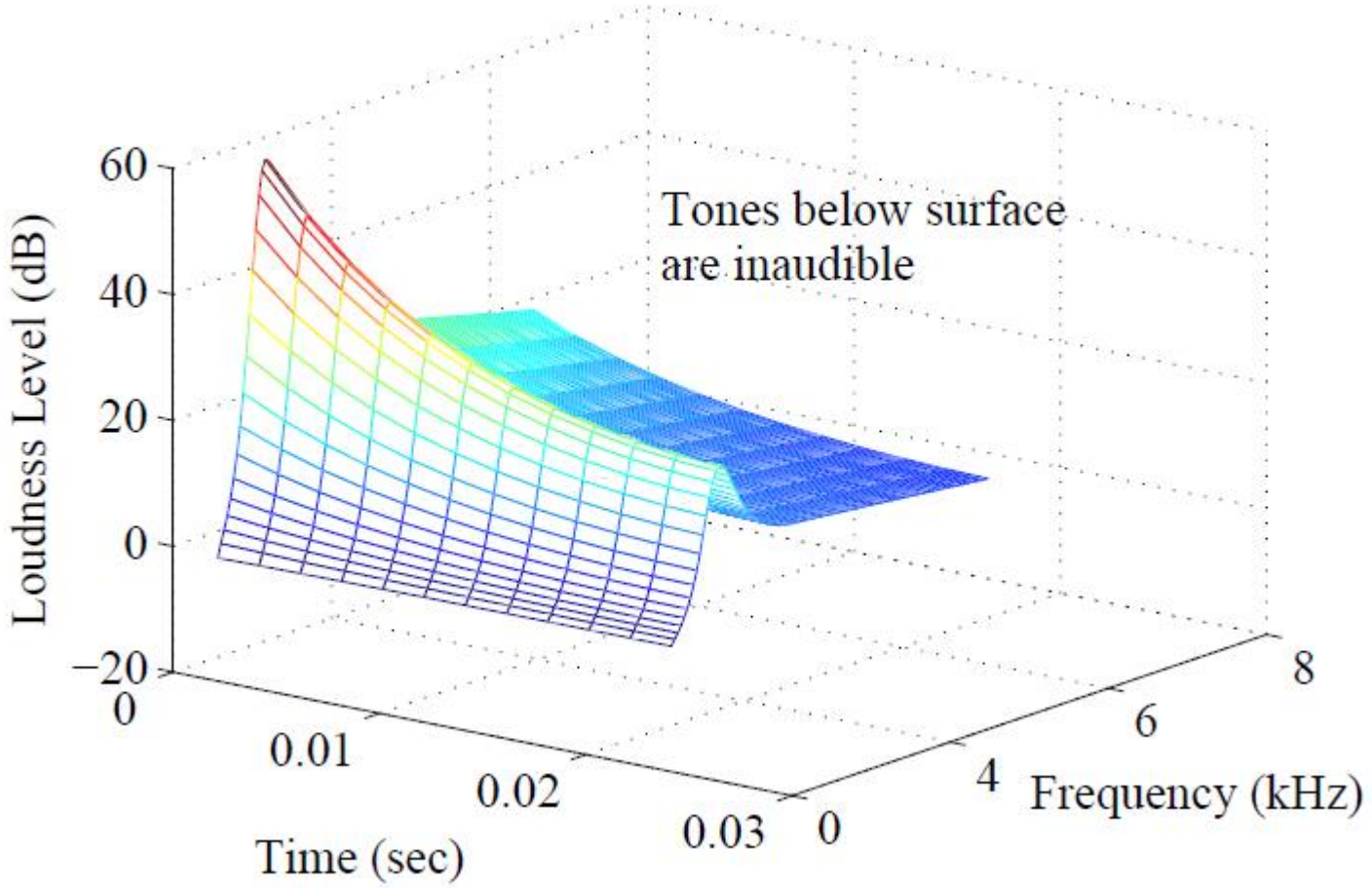


Fig. 14.7: Effect of temporal masking depends on both time and closeness in frequency.

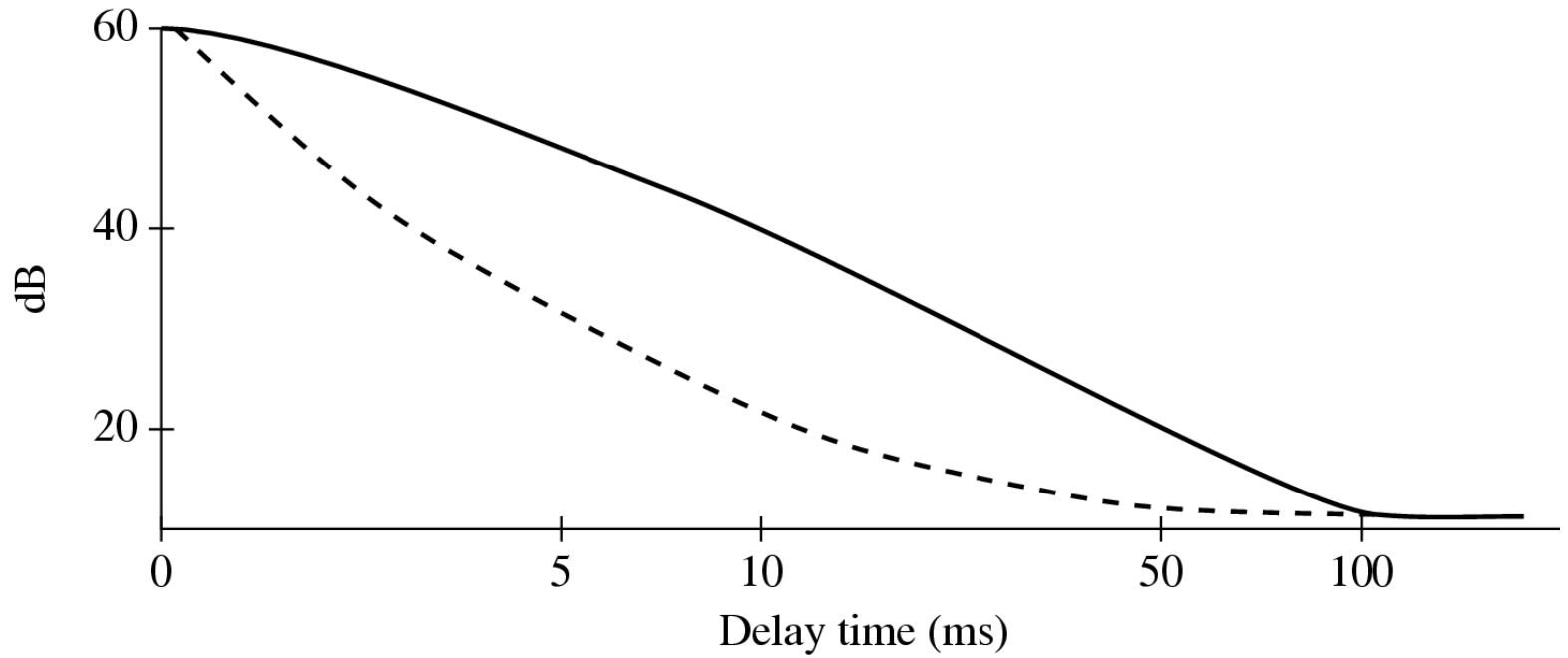


Fig.14.8: For a masking tone that is played for a longer time, it takes longer before a test tone can be heard. Solid curve: masking tone played for 200 msec; dashed curve: masking tone played for 100 msec.

14.2 MPEG Audio

- **MPEG audio compression** takes advantage of psychoacoustic models, constructing a large multi-dimensional lookup table to transmit masked frequency components using fewer bits.
- **MPEG Audio Overview**
  1. Applies a filter bank to the input to break it into its frequency components.
  2. In parallel, a psychoacoustic model is applied to the data for bit allocation block.
  3. The number of bits allocated are used to quantize the info from the filter bank – providing the compression.

MPEG Layers

- MPEG audio offers three compatible *layers*:
  - Each succeeding layer able to understand the lower layers.
  - Each succeeding layer offering more complexity in the psychoacoustic model and better compression for a given level of audio quality.
  - each succeeding layer, with increased compression effectiveness, accompanied by extra delay.
- The objective of MPEG layers: a good tradeoff between quality and bitrate.

MPEG Layers (Cont'd)

- Layer 1 quality can be quite good provided a comparatively high bitrate is available.
  - Digital Audio Tape typically uses Layer 1 at around 192 kbps.
- Layer 2 has more complexity; was proposed for use in Digital Audio Broadcasting.
- Layer 3 (MP3) is most complex, and was originally aimed at audio transmission over ISDN lines.
- Most of the complexity increase is at the encoder, not the decoder – accounting for the popularity of MP3 players.

MPEG Audio Strategy

- **MPEG approach to compression** relies on:
  - Quantization
  - Inaccuracy of human auditory system within the width of a critical band.

- **MPEG encoder** employs a bank of filters to:
  - Analyze the frequency ("spectral") components of the audio signal by calculating a frequency transform of a window of signal values.
  - Decompose the signal into subbands by using a bank of filters (Layer 1 & 2: "quadrature-mirror"; Layer 3: adds a DCT; psychoacoustic model: Fourier transform).

MPEG Audio Strategy (Cont'd)

- **Frequency masking**: by using a psychoacoustic model to estimate the just noticeable noise level:
  - Encoder balances the masking behavior and the available number of bits by discarding inaudible frequencies.
  - Scaling quantization according to the sound level that is left over, above masking levels.
- May take into account the actual width of the critical bands:
  - As mentioned earlier, audible frequencies are usually divided into 25 main critical bands (Table 14.1).
  - To keep simplicity, the MPEG model adopts a *uniform* width for all frequency analysis filters, using 32 overlapping subbands.

MPEG Audio Compression Algorithm

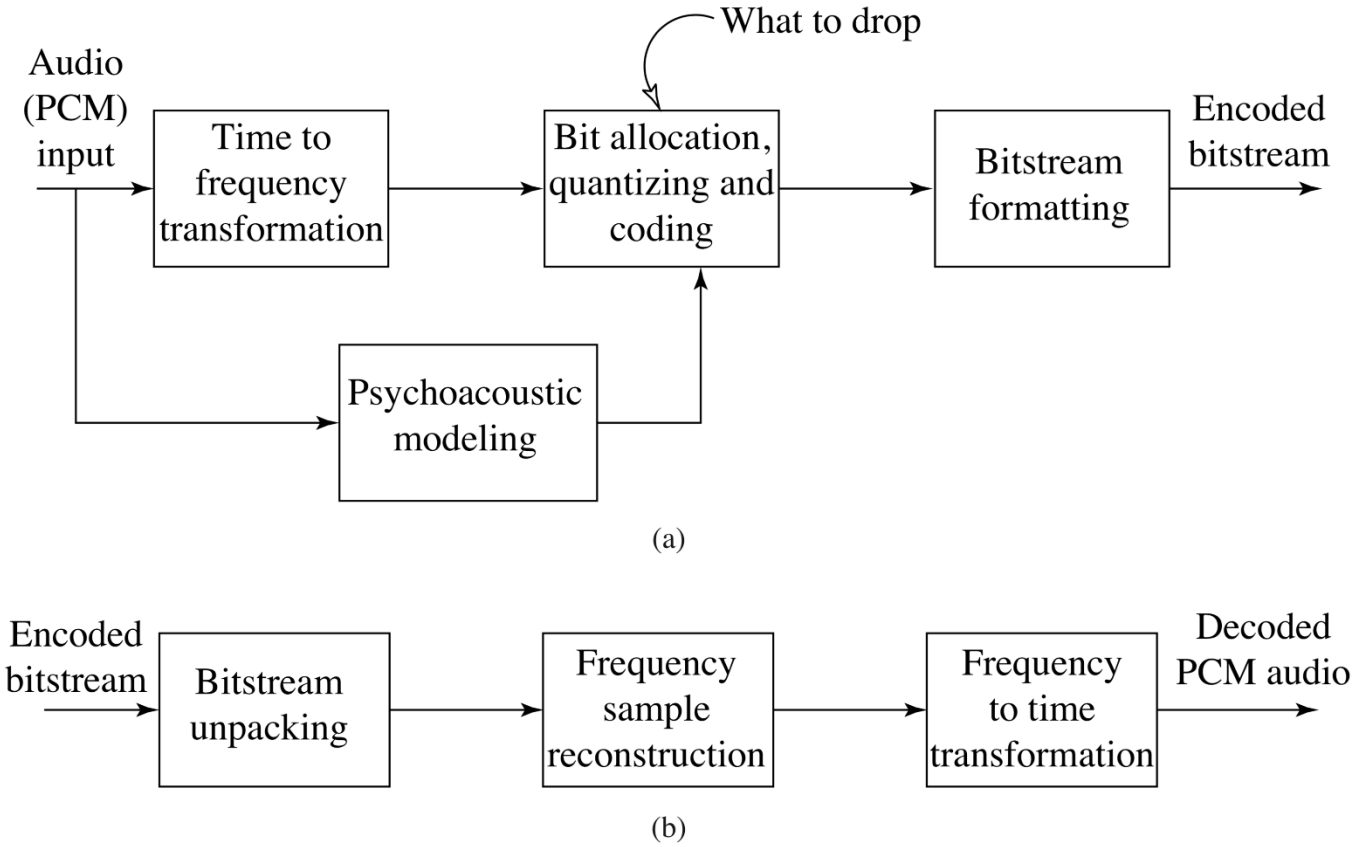


Fig. 14.9: Basic MPEG Audio encoder and decoder.

Basic Algorithm (Cont'd)

- The algorithm proceeds by dividing the input into 32 frequency subbands, via a filter bank.
  - A linear operation taking 32 PCM samples, sampled in time; output is 32 frequency coefficients.
- In the Layer 1 encoder, the sets of 32 PCM values are first assembled into a set of 12 groups of 32s.
  - an inherent time lag in the coder, equal to the time to accumulate 384 (i.e., 12×32) samples.
- Fig.14.11 shows how samples are organized.
  - A Layer 2 or Layer 3 frame actually accumulates more than 12 samples for each subband: a frame includes 1,152 samples.

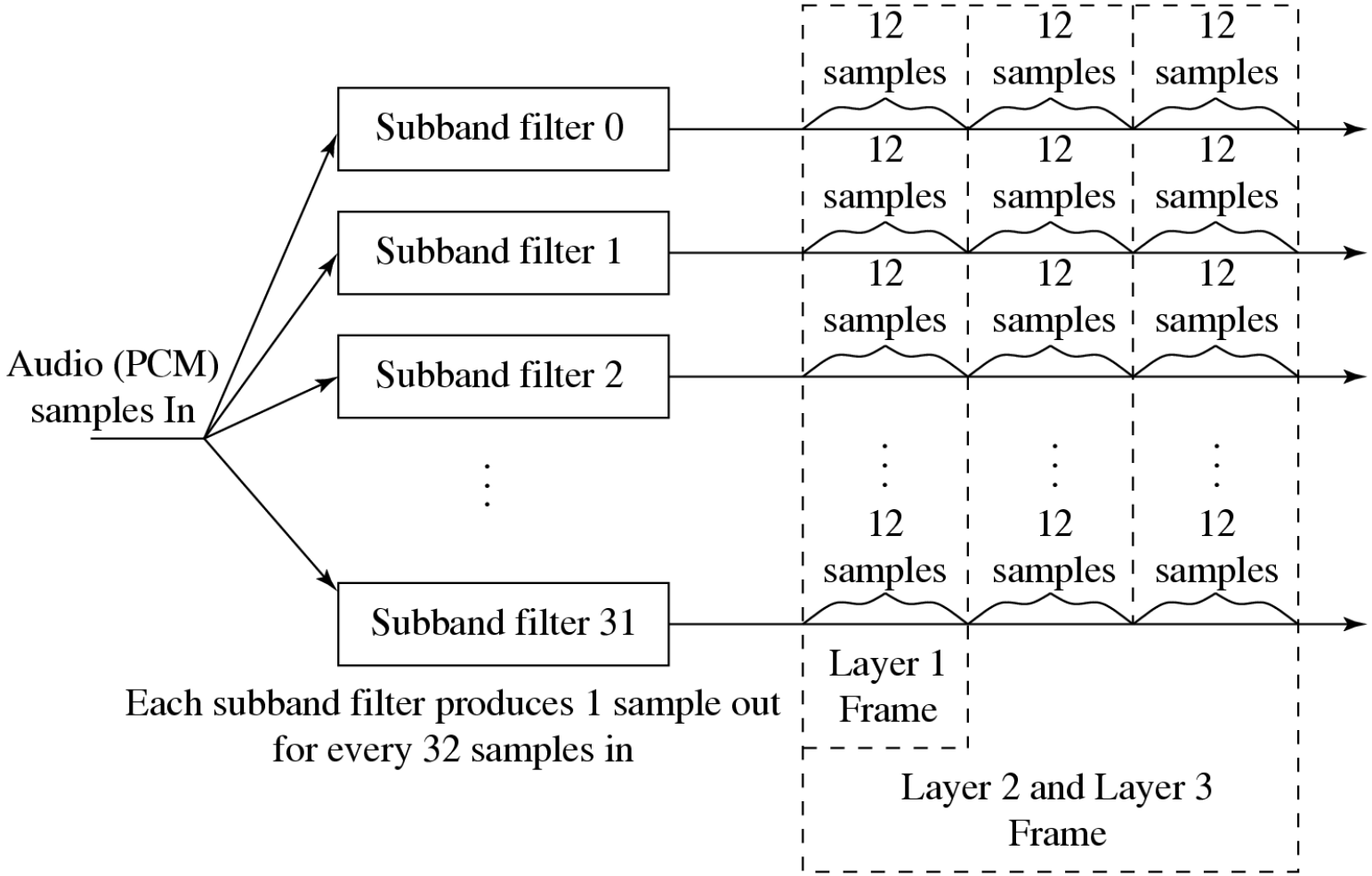


Fig. 14.11: MPEG Audio Frame Sizes

Bit Allocation Algorithm

- **Aim**: Ensure that all of the quantization noise is below the masking thresholds.
- **One common scheme**:
  - For each subband, the psychoacoustic model calculates the *Signal-to-Mask Ratio* (SMR) in dB.
  - Then the "Mask-to-Noise Ratio" (MNR) is defined as the difference (as shown in Fig.14.12):

$$\mathrm{MNR}_{\mathrm{dB}} \equiv \mathrm{SNR}_{\mathrm{dB}} - \mathrm{SMR}_{\mathrm{dB}} \tag{14.7}$$

  - Search for sub-band with the lowest MNR, and the number of code-bits allocated to this subband is incremented. (The sub-band with the lowest MNR has the highest priority for bit allocation)
  - Then a new estimate of the SNR is made, and the process iterates until there are no more bits to allocate, i.e., achieving the maximum number of bits.

SMR and MNR

- Signal to Mask Ratio, SMR, i.e. the difference in dB between the signal and the masking threshold; the latter is computed from the psychoacoustical model; a positive SMR value indicated that the signal is audible, while a negative one shows that the signal cannot be heard because masked by other signal components.
- Mask to Noise Ratio, MNR, i.e. the difference in dB between the masking value and the quantizing noise; its value has a very important qualitative meaning, since a negative value indicates that in that band the quantization noise is audible, while a positive one implies complete masking of the quantization noise.
  - in term of quantity, it indicates the dB margin for signal processing and how much the quantization noise can be heard.

Algorithm 14.1 (*Bit Allocation in MPEG Audio Compression (Layers 1 and 2)*)

1. From the psychoacoustic model, calculate the *Signal-to-Mask Ratio (SMR)* in decibels (dBs) for each subband:

$$SMR = 20 \log_{10} \frac{Signal}{Minimum\_masking\_threshold} \tag{14.6}$$

   - This determines the quantization, i.e., the minimum number of bits that is needed, if available. The amount of a signal above the threshold, i.e., SMR, s the amount that needs to be coded. Signals that are below the threshold do not.

2. Calculate *Signal-to-(quantization)-Noise Ratio (SNR)* for all signals.

   - A lookup table provides an estimate of SNR assuming a given number of quantizer levels.

3. *Mask-to-(quantization)-Noise Ratio (MNR)* is defined as the difference, in dB (See Fig. 14.12).

$$MNR = SNR - SMR \tag{14.7}$$

4. Iterate until no bits left to allocate

   - Allocate bits to the subband with the lowest MNR
   - Look up new estimate of SNR for the subband allocated more bits, and re-calculate MNR.

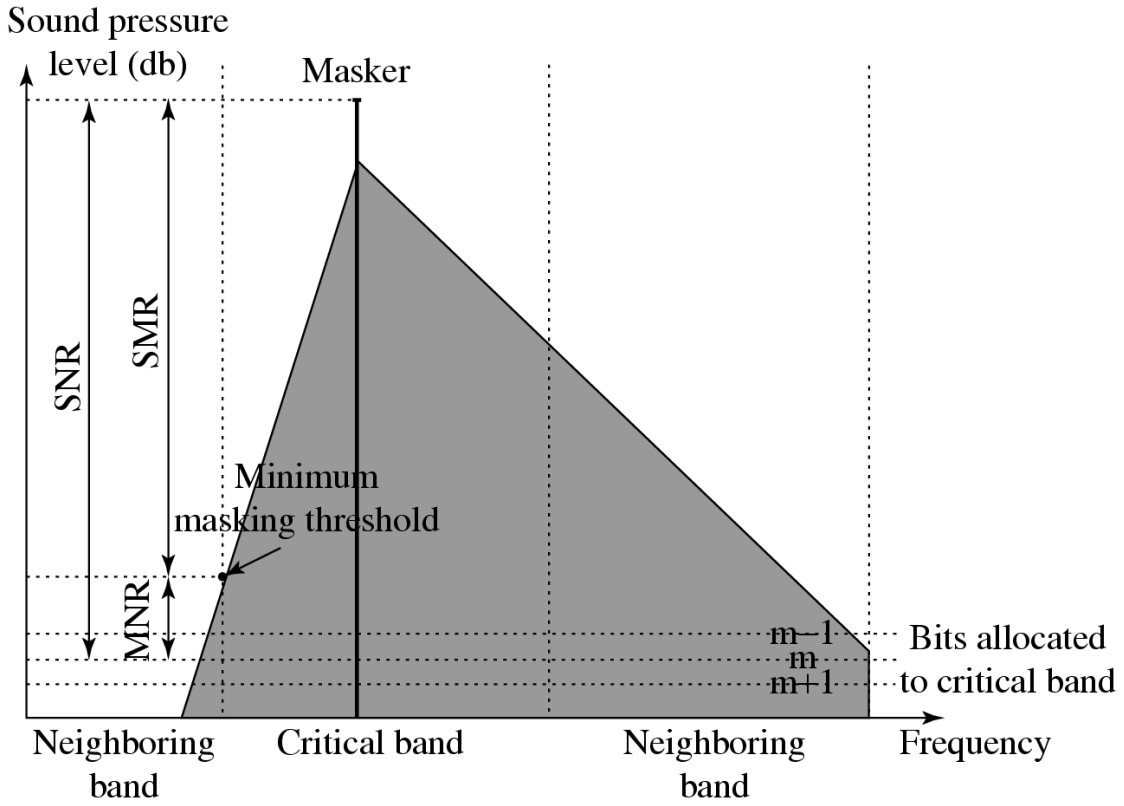


Fig. 14.12: MNR and SMR. A qualitative view of SNR, SMR and MNR are shown, with one dominate masker and m bits allocated to a particular critical band.

- Mask calculations are performed in parallel with subband filtering, as in Fig. 4.13:

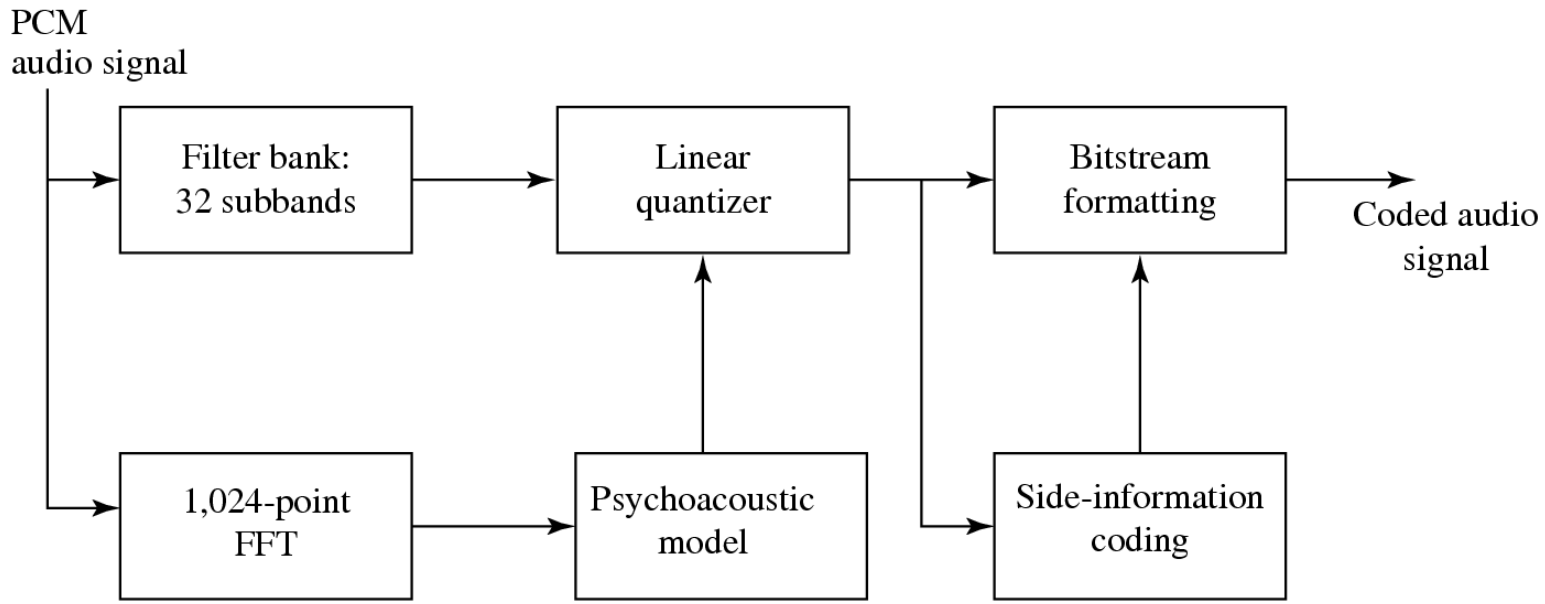


Fig. 14.13: MPEG-1 Audio Layers 1 and 2.

Layer 2 of MPEG-1 Audio

- **Main difference:**
  - Three groups of 12 samples are encoded in each frame and temporal masking is brought into play, as well as frequency masking.
  - Bit allocation is applied to window lengths of 36 samples instead of 12.
  - The resolution of the quantizers is increased from 15 bits to 16.
- **Advantage:**
  - a single scaling factor can be used for all three groups.

Layer 3 of MPEG-1 Audio

- **Main difference:**
  - Employs a similar filter bank to that used in Layer 2, except using a set of filters with non-equal frequencies.
  - Takes into account stereo redundancy.
  - Uses Modified Discrete Cosine Transform (MDCT) — addresses problems that the DCT has at boundaries of the window used by overlapping frames by 50%:

$$F(u)=2\sum_{i=0}^{N-1} f(i)\cos\left[\frac{2\pi}{N}\left(i+\frac{N/2+1}{2}\right)\left(u+1/2\right)\right], u=0,..,N/2-1 \quad (14.8)$$

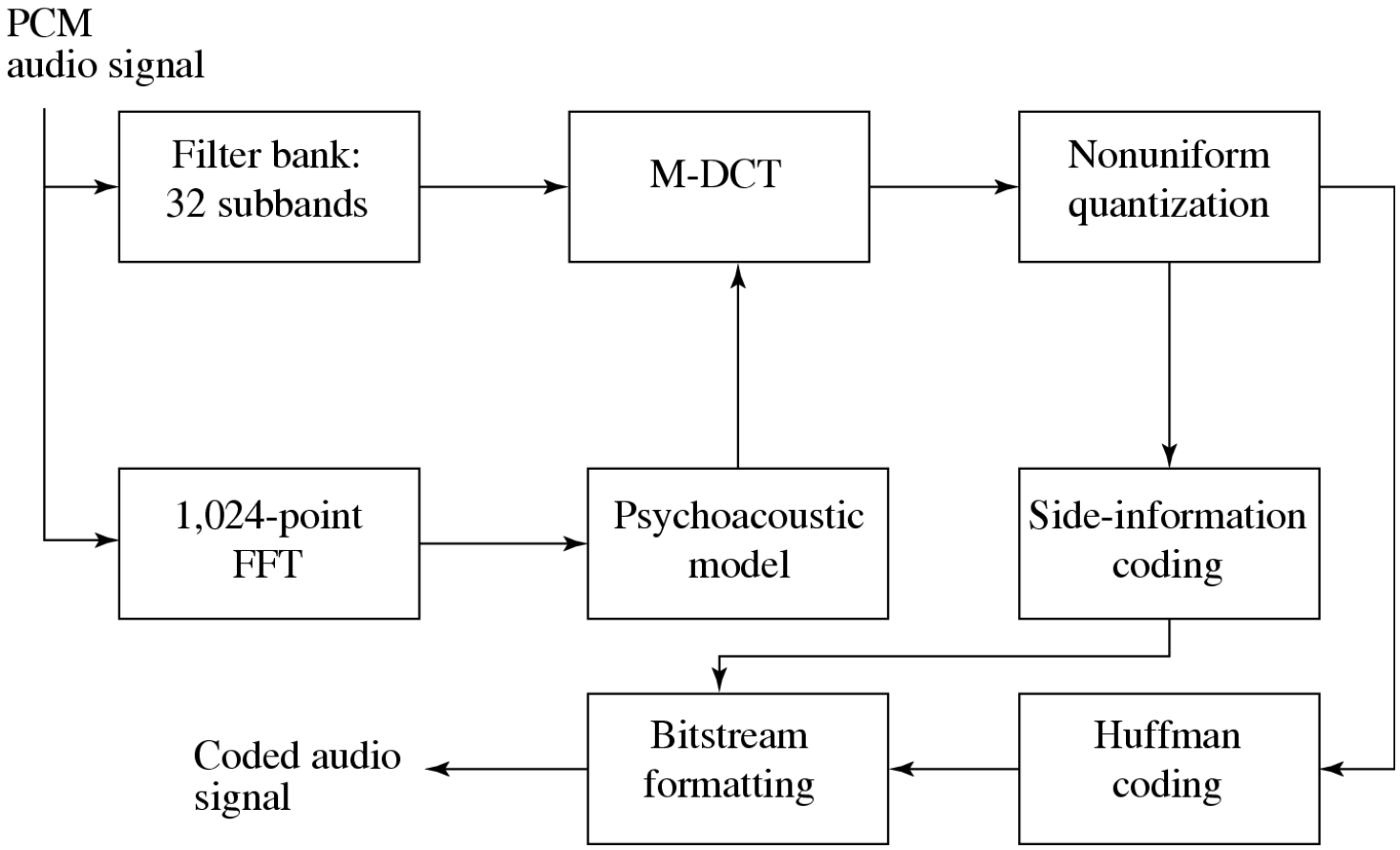


Fig 14.14: MPEG-Audio Layer 3 Coding.

- Table 14.2 shows various achievable MP3 compression ratios:

Table 14.2: MP3 compression performance

Sound Quality	Bandwidth	Mode	Compression Ratio
Telephony	3.0 kHz	Mono	96:1
Better than Short-wave	4.5 kHz	Mono	48:1
Better than AM radio	7.5 kHz	Mono	24:1
Similar to FM radio	11 kHz	Stereo	26 - 24:1
Near-CD	15 kHz	Stereo	16:1
CD	> 15 kHz	Stereo	14 - 12:1

MPEG-2 AAC (Advanced Audio Coding)

- AAC was developed to succeed MP3 for digital audio; delivers better sound quality than MP3 for the same bitrate.
- Default audio format for YouTube, iPhone, iTunes, Nintendo, and PlayStation.
- The standard vehicle for DVDs:
  - Audio coding technology for the DVD-Audio Recordable (DVD-AR) format, also adopted by XM Radio.
- Aimed at transparent sound reproduction for theaters.
  - Can deliver this at 320 kbps for five channels so that sound can be played from 5 different directions: Left, Right, Center, Left-Surround, and Right-Surround.
  - 5.1 channel systems also include a low-frequency enhancement (LFE) channel (a “woofer”).
- Also capable of delivering high-quality stereo sound at bit-rates below 128 kbps.

MPEG-2 AAC (Cont'd)

- Support up to 48 channels, sampling rates between 8 kHz and 96 kHz, and bit-rates up to 576 kbps per channel.
- Like MPEG-1, MPEG-2, supports three different "profiles", but with a different purpose:
  - *Main* profile
  - *Low Complexity*(LC) profile
  - *Scalable Sampling Rate* (SSR) profile

MPEG-4 Audio

- Integrates several different audio components into one standard: speech compression, perceptually based coders, text-to-speech, and MIDI.
- *MPEG-4 AAC* (*Advanced Audio Coding*), is similar to the MPEG-2 AAC standard, with some minor changes.
- **Perceptual Coders**
  - Incorporate a *Perceptual Noise Substitution* module.
  - Include a *Bit-Sliced Arithmetic Coding* (BSAC) module.
  - Also include a second perceptual audio coder, a vector-quantization method entitled TwinVQ.

MPEG-4 Audio (Cont'd)

- **Structured Coders**
  - Takes "Synthetic/Natural Hybrid Coding" (SNHC) in order to have very low bit-rate delivery an option.
  - **Objective**: integrate both "natural" multimedia sequences, both video and audio, with those arising synthetically – "structured" audio.
  - Takes a "toolbox" approach and allows specification of many such models.
  - E.g., *Text-To-Speech* (TTS) is an ultra-low bit-rate method, and actually works, provided one need not care what the speaker actually sounds like.

14.3 Other Audio Codecs

- Ogg Vorbis is an open-source audio compression format.

Table 14.3: Comparison of MP3, MPEG-4 AAC, and Ogg Vorbis

	MP3	MPEG-4 AAC	Ogg vorbis
File extension	.mp3	.aac, .mp4, .3gp	.ogg
Original name	MPEG-1 Audio Layer 3	Advanced Audio Coding	Ogg
Developer	CCETT, IRT, Fraunhofer Society	Fraunhofer IIS, AT&T Bell Labs, Dolby, Sony Corp., and Nokia	Xiph.org Foundation
Released	1994	1997	v1.0 frozen May 2000
Algorithm	lossy compression	lossy compression	lossy compression
Quality	Lower quality than AAC and Ogg	Better quality at same bit rate as MP3	Better quality and smaller file size than MP3 at same bit rates
Used in	Default standard for audio files	iTunes raised its popularity	Open-source platform

14.4 MPEG-7 Audio and Beyond

- Table 14.4 summarizes the target bitrate ranges and main features of other modern general audio codecs.

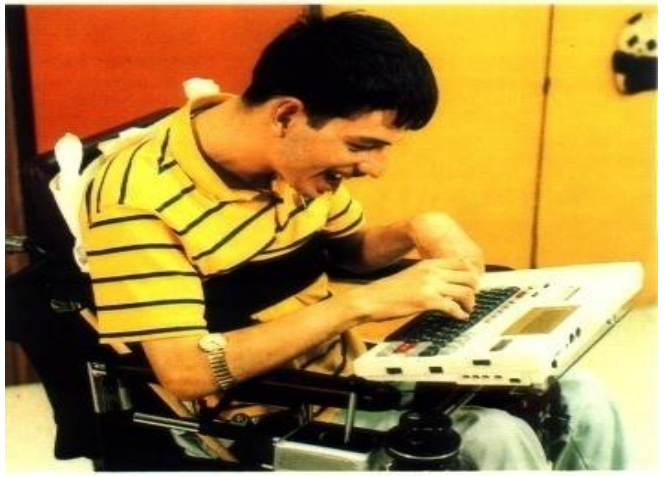
Table 14.4: Comparison of audio coding systems

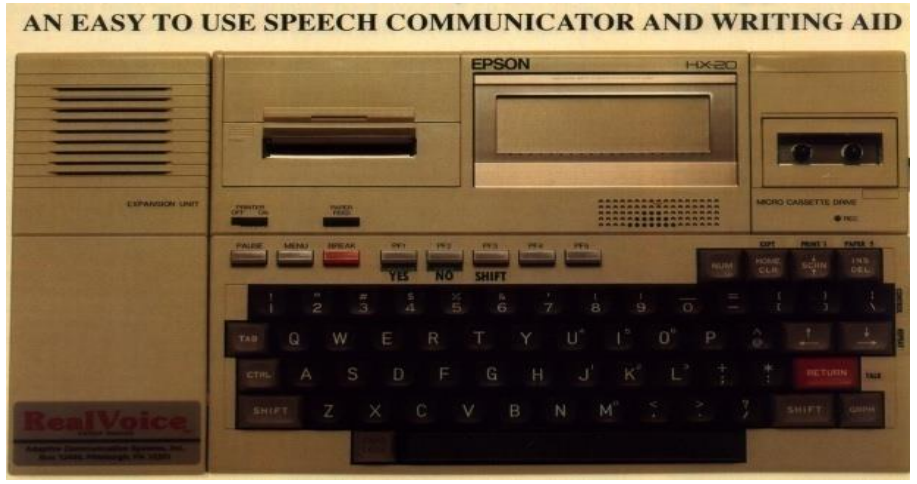
Codec	Bitrate kbps/channel	Complexity	Main Application
Dolby AC-2	128–192	Low (encoder/decoder)	Point-to-point, cable
Dolby AC-3	32–640	Low (decoder)	HDTV, cable, DVD
Dolby Digital Plus (Enhanced AC-3)	32–6,144	Low (decoder)	HDTV, cable, DVD
DTS: Digital Surround	8–512	Low (for lossless audio extension)	DVD, entertainment, professional
WMA: Windows Media Audio	128–768	Low (low-bit-rate streaming)	Many applications
MPEG SAOC	As low as 48	Low (decoder/rendering)	Many applications

14.4 MPEG-7 Audio and Beyond

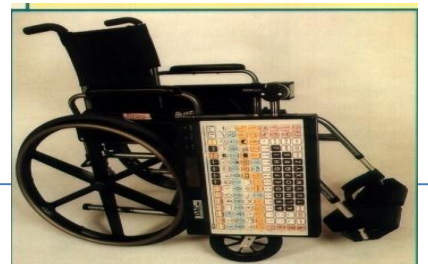
- **Difference** from current standards:
- MPEG-4 is aimed at compression using objects.
- MPEG-7 is mainly aimed at "search": How can we find objects, assuming that multimedia is indeed coded in terms of objects.
  - The objective, in terms of audio, is to facilitate the representation and search for sound content, perhaps through the tune or other descriptors.
  - An example application supported by MPEG-7 is automatic speech recognition (ASR).
  - Further standards in the MPEG sequence are mostly not aimed at further audio compression standardization. For example, MPEG-DASH (Dynamic Adaptive Streaming over HTTP) is aimed at streaming of multimedia.

聽語障輔助




AN EASY TO USE SPEECH COMMUNICATOR AND WRITING AID
EPSON
HX-20
RealVoice


DAC36
DAC — DIGITAL AUGMENTATIVE COMMUNICATOR
TALK
DELETE
JOIN
SHIFT
SPACE



聽力喪失

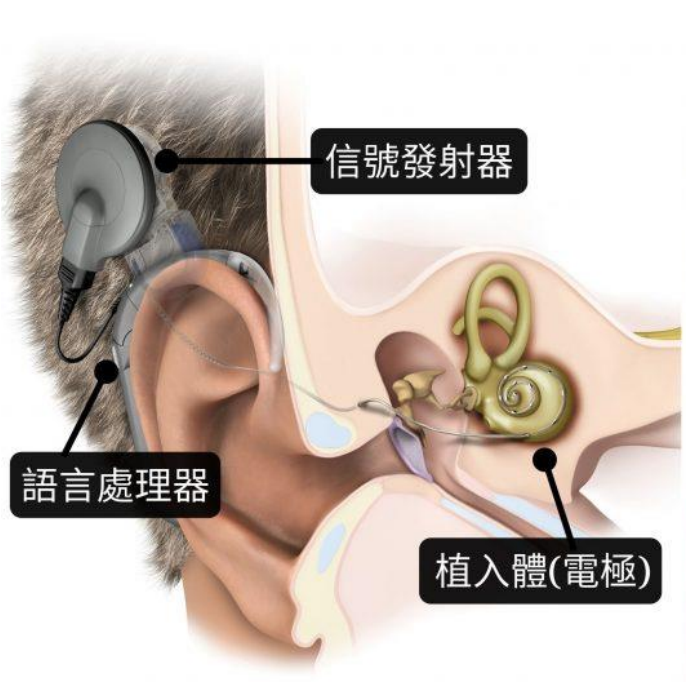
- 聽力喪失主要原因有2種:感覺性聽力喪失及傳導性聽力喪失。
- 傳導性聽力喪失起因於中耳或耳道的機械性問題，阻斷聲音的傳導，這類聽力喪失通常可以經由藥物、手術或助聽器復原及改善
- 感覺性聽力喪失因內耳、聽神經或腦部神經路徑受損，多數可以經由助聽器獲得助益。
- 少數感覺性聽損較嚴重者，傳統助聽器無法獲得良好的語音聽辯力，利用人工電子耳才可能大幅改善患者聽力。

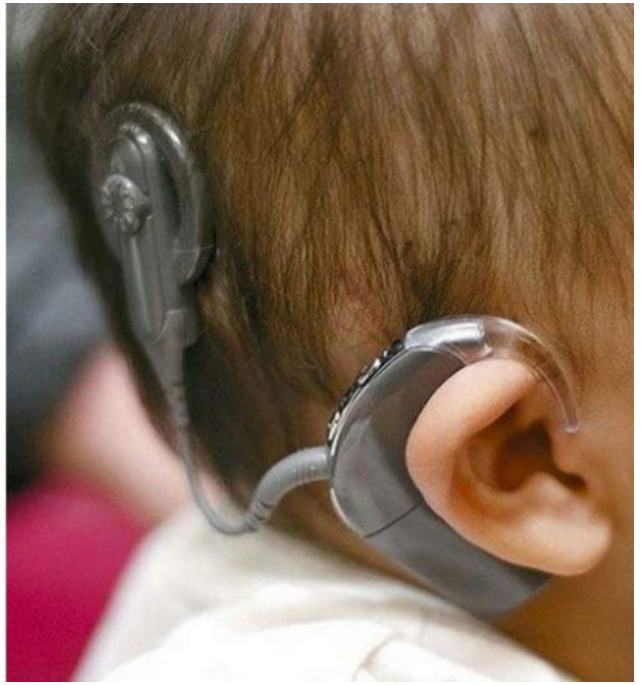
助聽器

麥克風接收外部聲波後，會將聲音轉換成電波，經擴大器/晶片根據助聽器本身程式編碼轉成使用者本身所需要的強度根據調整程度選擇對應喇叭功率，接收器再將對應的功率電能轉回聲波，傳入配戴者的耳中

人工電子耳

人工電子耳分為耳內、耳外兩部分，耳外部分外觀近似於一般助聽器，透過掛在耳殼上的「語言處理器」，將聲音轉換為電流訊號，接著由吸附在頭皮表面的「信號發射器」與頭皮底下的線圈感應；耳內植入體則以手術方式埋藏頭皮下，由線圈連接細長的電極進入耳蝸





25 Critical Bands and Their Bandwidths

Band #	Lower Bound (Hz)	Center (Hz)	Upper Bound (Hz)	Bandwidth (Hz)
1	-	50	100	-
2	100	150	200	100
3	200	250	300	100
4	300	350	400	100
5	400	450	510	110
6	510	570	630	120
7	630	700	770	140
8	770	840	920	150
9	920	1000	1080	160
10	1080	1170	1270	190
11	1270	1370	1480	210
12	1480	1600	1720	240

Band #	Lower Bound (Hz)	Center (Hz)	Upper Bound (Hz)	Bandwidth (Hz)
13	1720	1850	2000	280
14	2000	2150	2320	320
15	2320	2500	2700	380
16	2700	2900	3150	450
17	3150	3400	3700	550
18	3700	4000	4400	700
19	4400	4800	5300	900
20	5300	5800	6400	1100
21	6400	7000	7700	1300
22	7700	8500	9500	1800
23	9500	10500	12000	2500
24	12000	13500	15500	3500
25	15500	18775	22050	6550

聽障者的溝通方式

- **手語**
  - 手語是一種視覺性語言，是聽障者學習輔助工具
  - 它的表達方式是將所有語音化為手勢，再加上面部表情或肢體動作。
- **口語**
  - 聽障者利用殘存聽力，運用讀話技術接受外來的訊息，再利用視覺觸覺及殘存聽力學習說話表達自己。
- **綜合溝通法**
  - 利用聲音、唇形及手語將訊息傳給對方，因為在溝通的過程中，聽覺、視覺是相輔相成的。
- **筆談**
  - 筆談的方法，除使用紙筆書寫外，尚可採取手指在空中拼字或寫字「空書」，以及用手指在掌心寫字的「畫掌法」等型式。

聽障者手語輔助

美國17歲男孩發明手語翻譯手套

這種“手語翻譯手套”內裝有多個感測器，帶有一個小型的信號傳送器，一個獨立的、能裝入口袋的接收器。使用時，接收器能將聽障者的手勢所隱含的意思以文字形式表達出來，自動顯示在螢幕上。由於聽障者每天要接觸很多不懂手語的正常人，如何讓正常人理解他們心中的想法，是多數聽障者人在溝通中遇到的難題。

範例—國外

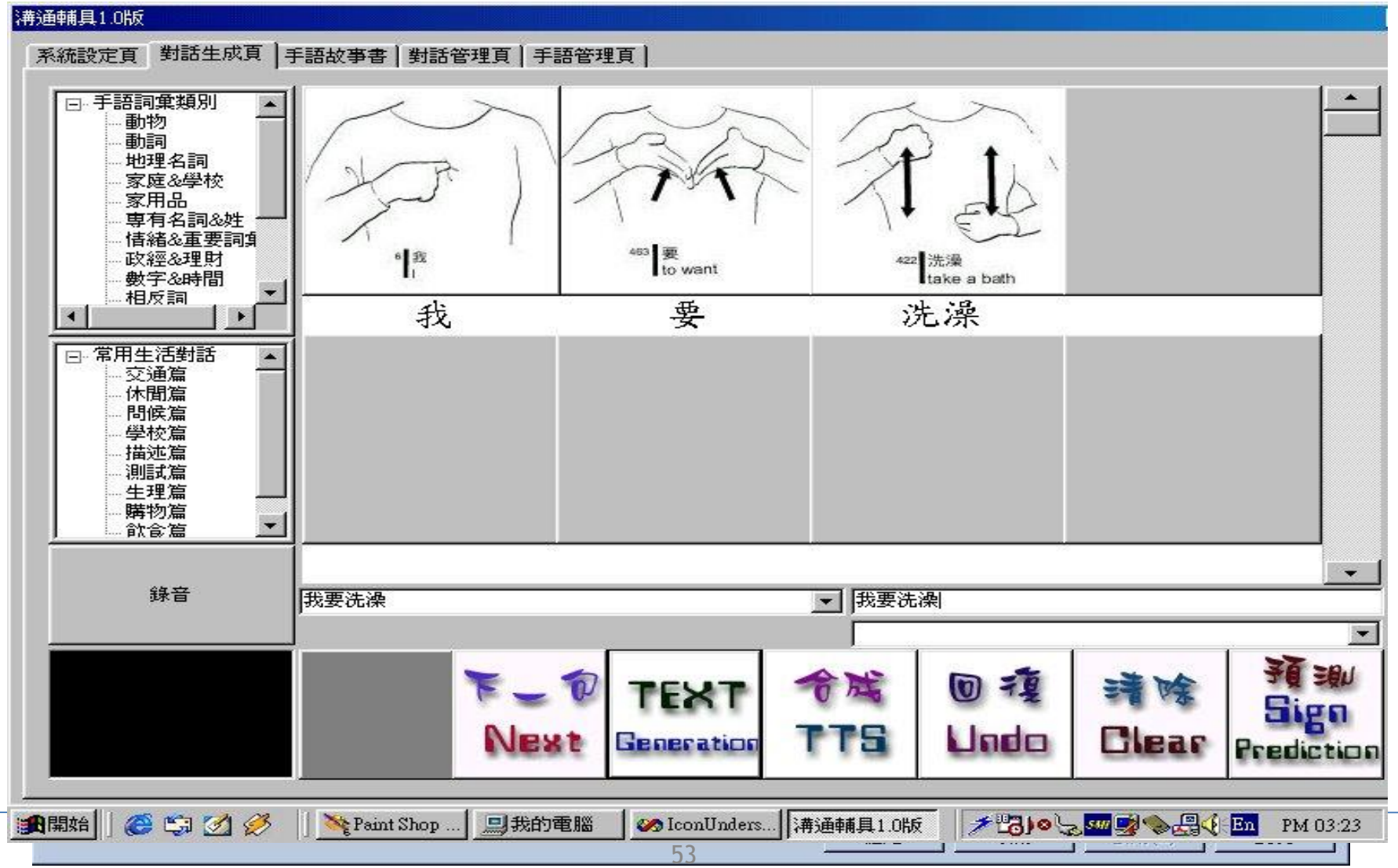
美國 MIT Media Lab：手語辨識系統

美國MIT之Media Lab的Starner、Weaver與Pentland三位教授研發一套可戴式的手語辨識系統，其裝置係以一頂棒球帽承載一具相機，將視訊傳至一具迷你電腦，以進行美國手語辨識，目前辨識成功率可達97%。

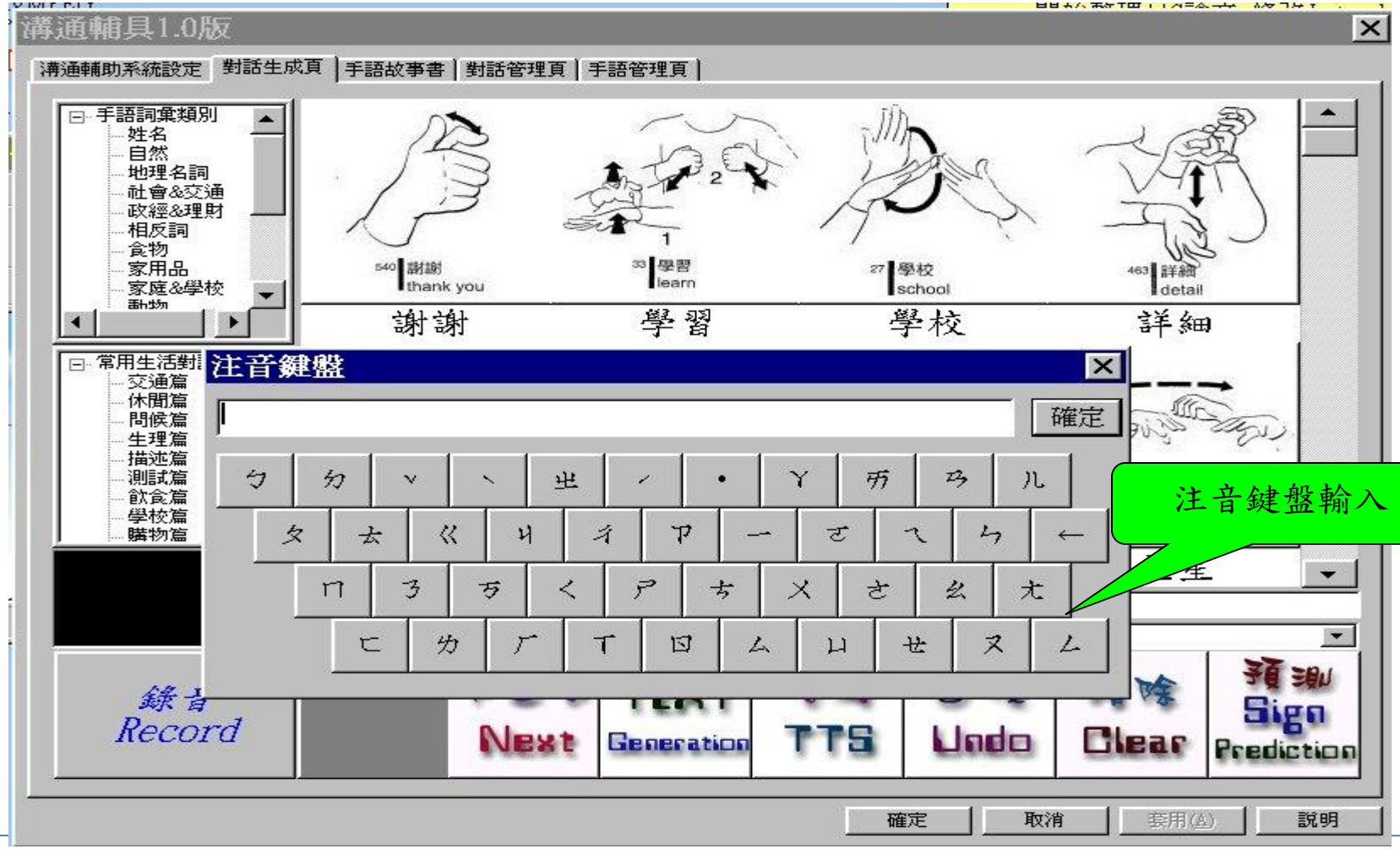
日本 JAIST：SYUWAN 手語翻譯系統

日本北陸先端科學技術大學院大學 (Japan advanced institute of science and technology, JAIST) 之情報科學研究科 (school of information science)的 Tokuda 與Okumura研發一套稱之為 SYUWAN的日本手語翻譯系統，目前成功率已達70%。

手語輸入:手語鍵盤輸入


溝通輔具1.0版
系統設定頁
對話生成頁
手語故事書
對話管理頁
手語管理頁
手語詞彙類別
動物
動詞
地理名詞
家庭&學校
家用品
專有名詞&姓
情緒&重要詞
政經&理財
數字&時間
相反詞
常用生活對話
交通篇
休閒篇
問候篇
學校篇
描述篇
測試篇
生理篇
購物篇
飲食篇
我
I
要
to want
洗澡
take a bath
我
要
洗澡
錄音
我要洗澡
我要洗澡
Next
TEXT
Generation
合成
TTS
回復
Undo
清除
Clear
預測
Sign
Prediction
開始
Paint Shop ...
我的電腦
IconUnders...
溝通輔具1.0版
En
PM 03:23

手語輸入:注音簡易輸入

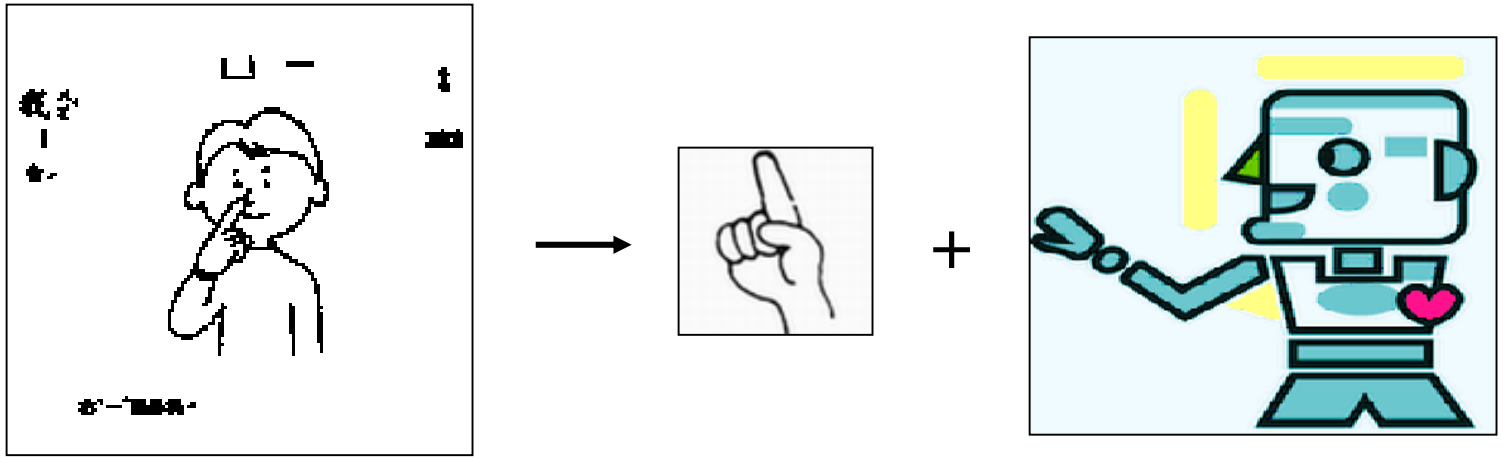

溝通輔具1.0版
溝通輔助系統設定
對話生成頁
手語故事書
對話管理頁
手語管理頁
手語詞彙類別
姓名
自然
地理名詞
社會&交通
政經&理財
相反詞
食物
家用品
家庭&學校
謝謝
學習
學校
詳細
注音鍵盤
確定
注音鍵盤輸入
Record
Next
Generation
TTS
Undo
Clear
Sign
Prediction
確定
取消
說明

手語輸入:手語手勢碼輸入

運用基本手勢碼，達到檢索手語詞彙之目的

舉例:

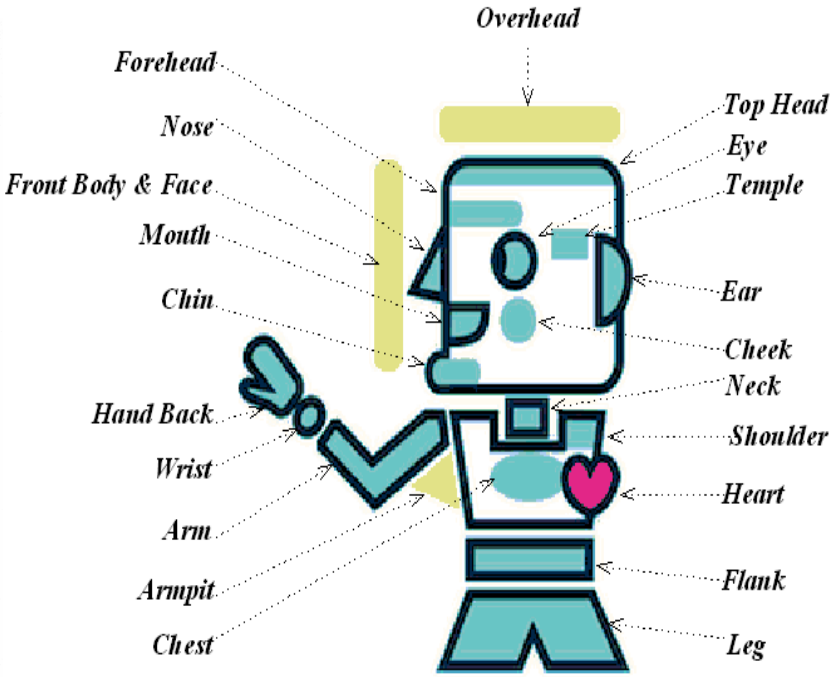
『我』 = 手形「食指」 + 位置「鼻子」



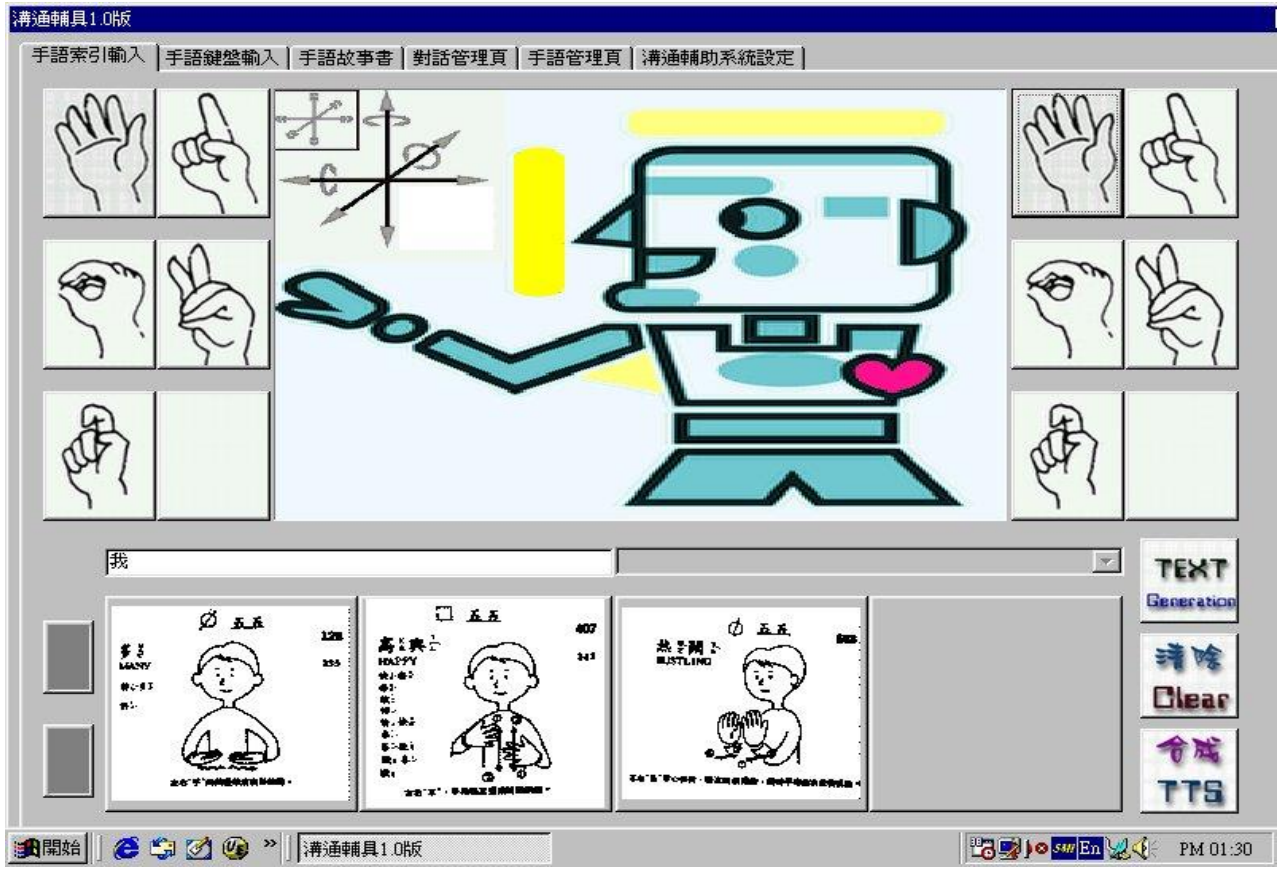
手語手勢碼

Hand Shapes (dez) & Locations (tab)

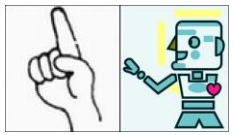
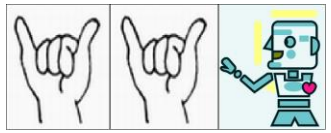
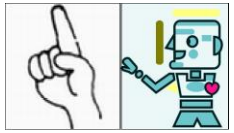
HS1	HS2	HS3	HS4	HS5	HS6	HS7	HS8	HS9
Zero	One	Two	Three	Four	Five	Six	Seven	Eight
HS10	HS11	HS12	HS13	HS14	HS15	HS16	HS17	HS18
Nine	Ten	Twenty	Thirty	Forty	Eighty	Hundred	Thousand	Ten Housand
HS19	HS20	HS21	HS22	HS23	HS24	HS25	HS26	HS27
Female	Hand	Rectangle	Sentence	Brother	People	Together	Keep	Male
HS28	HS29	HS30	HS31	HS32	HS33	HS34	HS35	HS36
Lu	Sister	Tiger	Fruit	Hu	Very	Airplane	Zhi	Fist
HS37	HS38	HS39	HS40	HS41	HS42	HS43	HS44	HS45
Borrow	Gentle	Secondary	Brown	Child	Vegetable	Pen	Like	Duck
HS46	HS47	HS48	HS49	HS50	HS51			
Money	Dragon	Worn	Arm	Difficult	WC			



手語手勢碼輸入


溝通輔具1.0版
手語索引輸入
手語鍵盤輸入
手語故事書
對話管理頁
手語管理頁
溝通輔助系統設定
我
TEXT
Generation
清除
Clear
合成
TTS
開始
溝通輔具1.0版
PM 01:30

自動中文文句生成


不可以	西瓜	一	吃
我	西班牙	一次	蔥
奇怪		一個禮拜	食
蟑螂		丁	零食
		人	點心
		刀子	
		小	

候選關鍵詞網格(Word Lattice)問題：

- 搜尋空間龐大
- 存在轉譯問題
- 如何選出最佳轉譯生成之語句

比手話語
TaiSign Communicator

中文轉譯手語

詞序

相對於中文有文法/句型轉置

- 虛詞省略：非常漂亮的狗→ 狗/漂亮/非常
- 數量詞省略：我有兩個姊姊 → 我/姊姊/兩/有
- 否定詞置後：他沒有告訴我→他/告訴/我/沒有

詞彙對應

多對少

- 漂亮 →美麗;　快樂→高興;　難過→悲傷..

複合詞

- 認識:臉+知道;　專有名詞(地名、人名):特徵、空書…

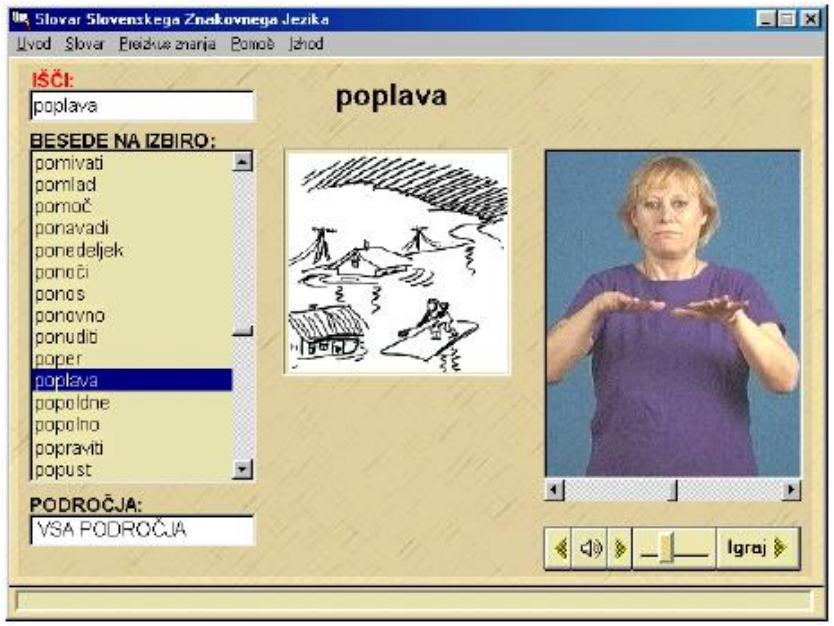
國外範例

Slovenia Sign Language (F. Solina, et al.)

手部位置考量串接個別手語影像

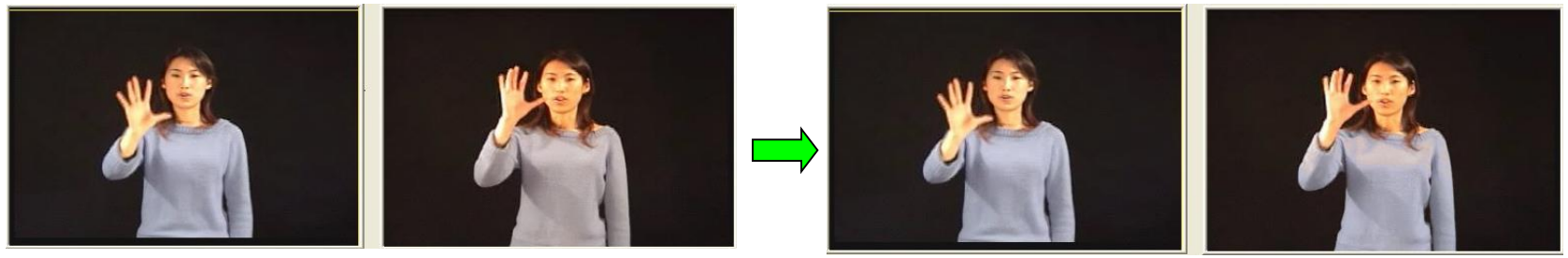
German Sign Language (R. Kennaway, 2000)

虛擬人物與標記式符號描述手語動作



影帶校正及標記

- 顏色校正

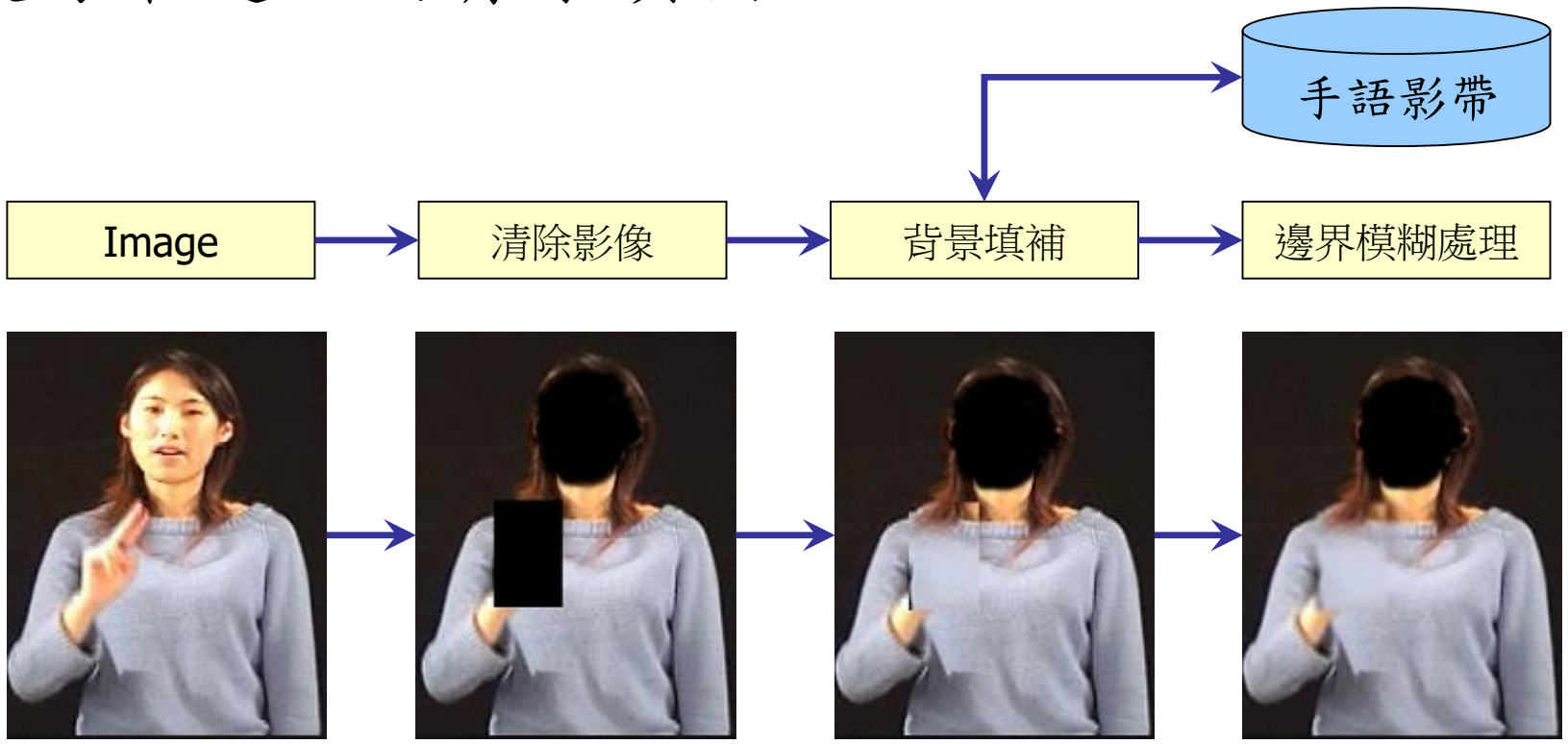


- 位置校正

- 手型及位置標記

臉部及手形清除

對影帶搜尋最適合的背景填補

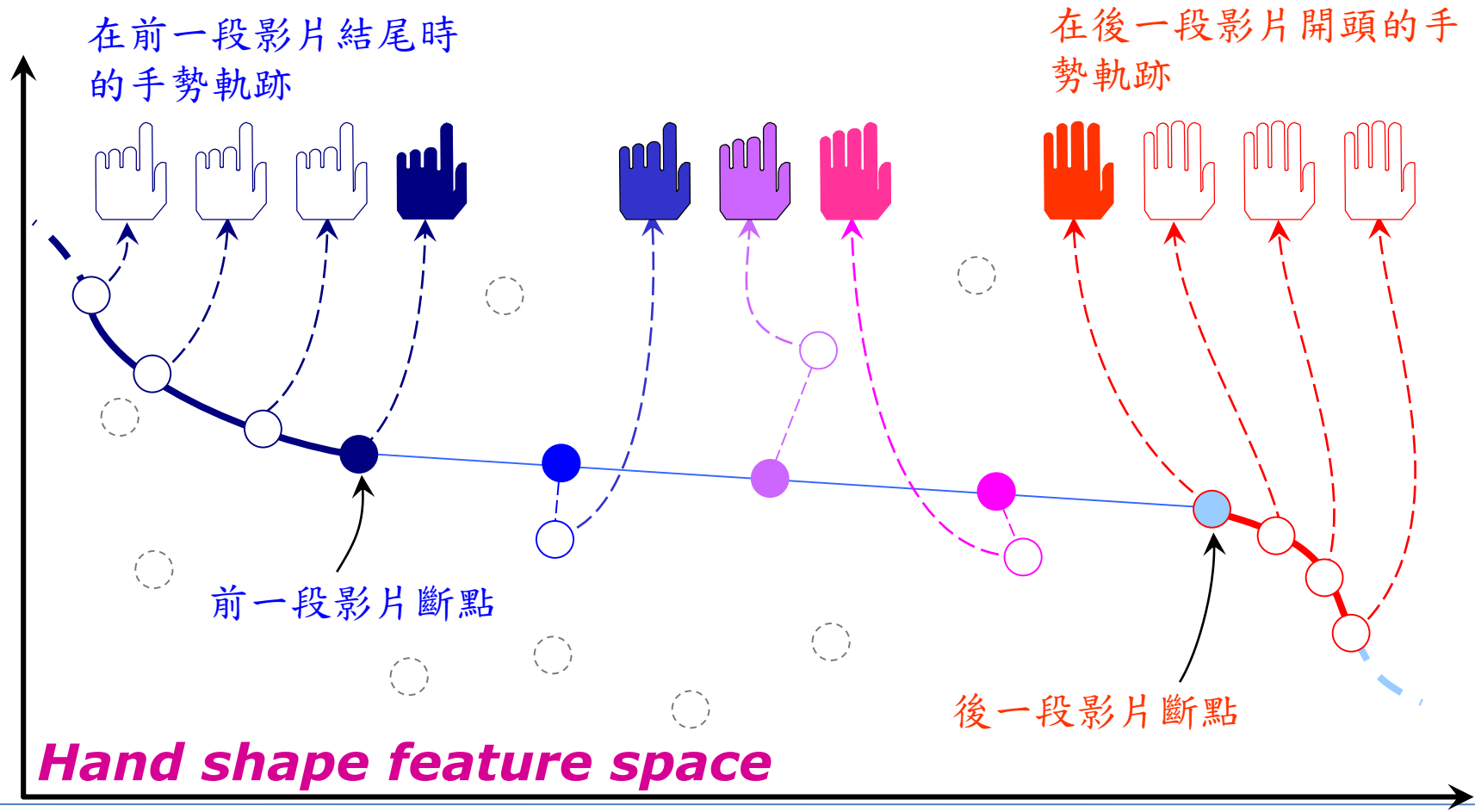


臉部影像內插合成

用兩兩串接的臉部影像作內差合成

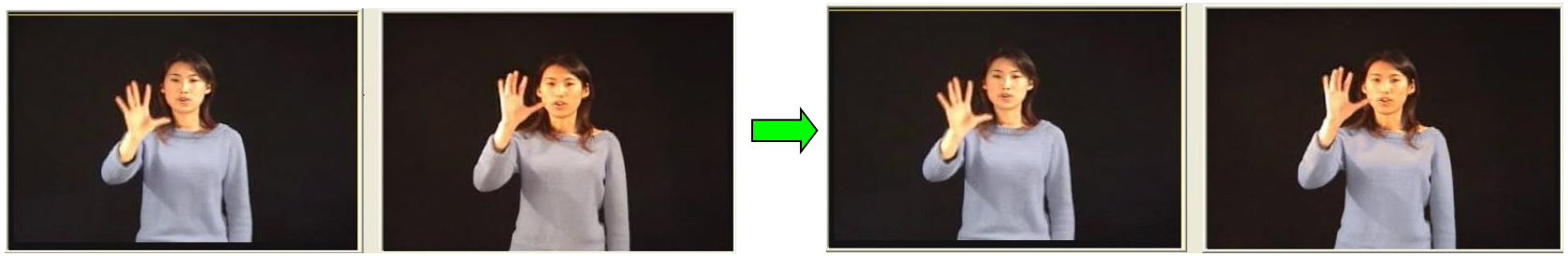
平滑化串接影像的臉部表情

手形內插串接

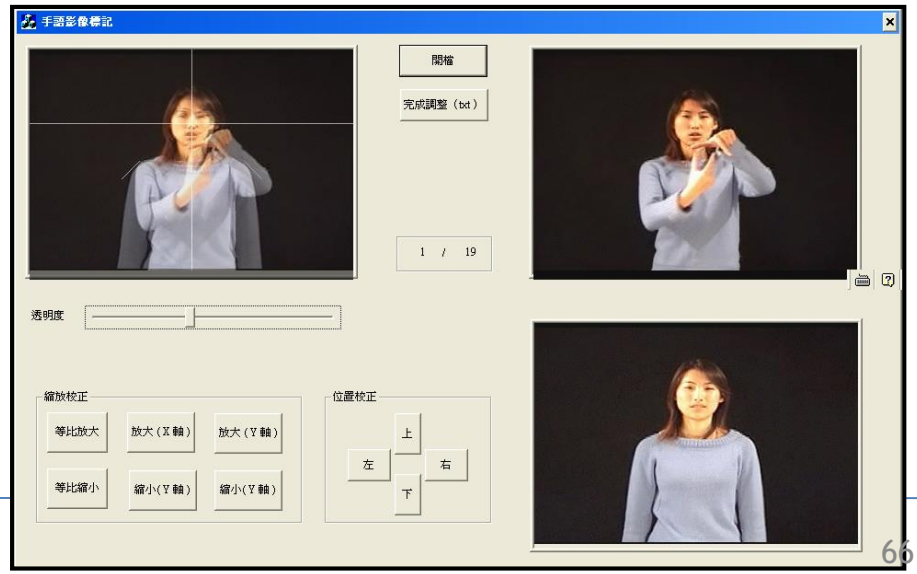

在前一段影片結尾時
的手勢軌跡
在後一段影片開頭的手
勢軌跡
前一段影片斷點
後一段影片斷點
Hand shape feature space

Video PreProcessing

- Color Normalization



- Position calibration



- Handshape and Position Tagging

Chinese to Sign Language Translation

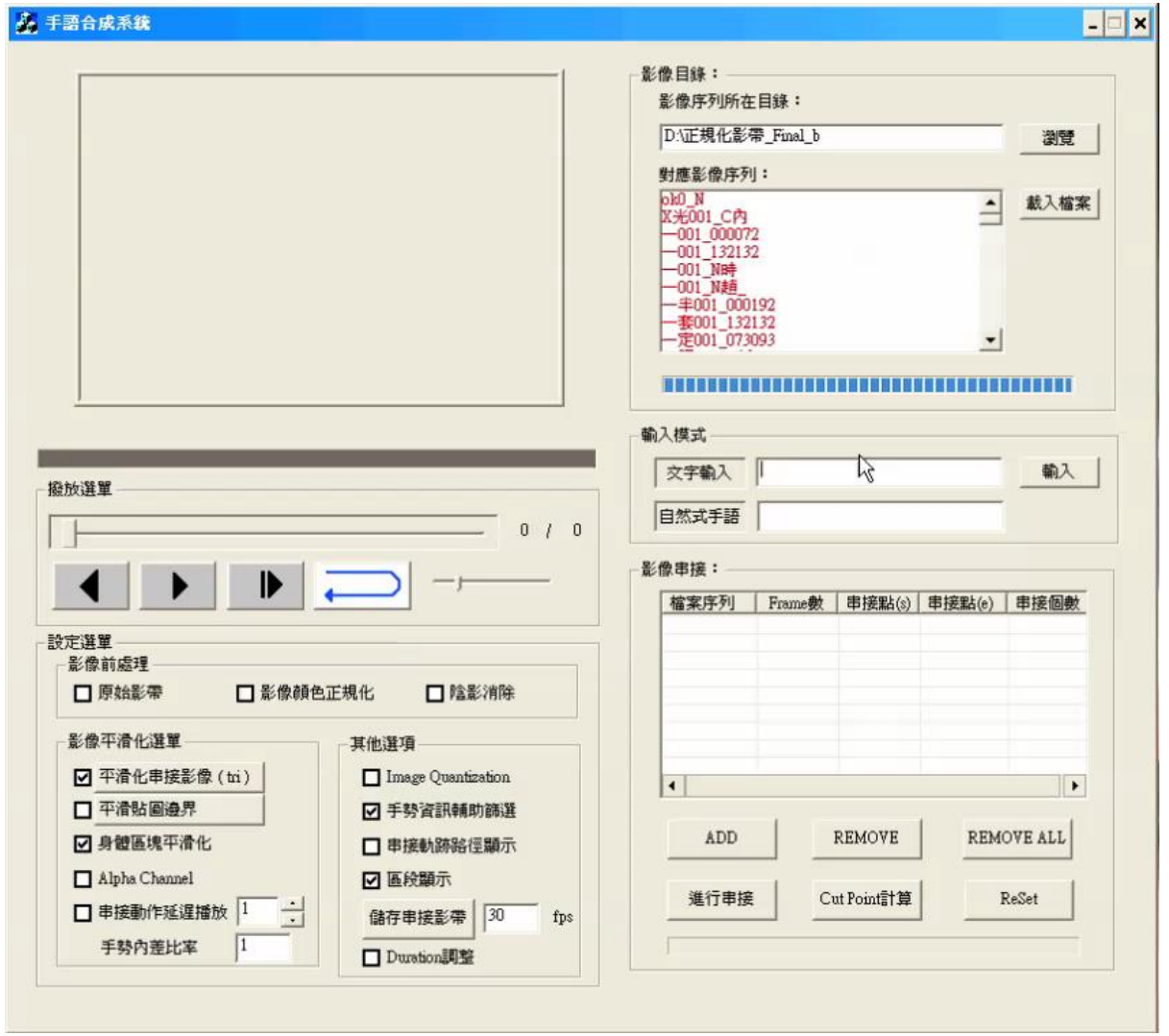
Chinese:

我去泰國

(I want to go to Thailand)

TSL:

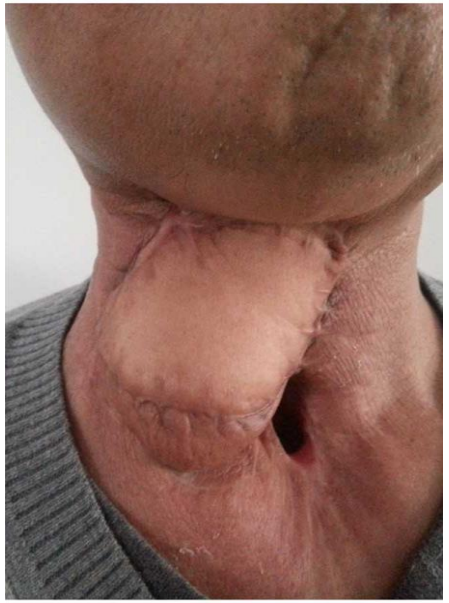
我(I) 泰國(Thailand) 去(go)



全喉切除病人溝通

晚期喉癌患者因腫瘤廣泛侵犯，故需合併全喉切除術以減少癌症復發機會。

常見的輔助工具是氣動式人工發聲器，另外也有電動人工發聲器、人工發聲瓣等。部份病人也可以透過復健學習食道語，而不需任何輔助工具。



Electrolarynx (EL) speech

- The Electrolarynx (EL) speech is typically spoken with an EL device which generates excitation signals to substitute human vocal fold vibrations
- The EL device offers the possibility to re-obtain speech when the larynx is removed after a total laryngectomy.
- EL speech is under–resourced data:
  - Recording EL speech is difficult
  - Existing EL database is difficult to obtain

Electrolaryngeal speech (Throat EL speech)

Electrolarynx Speech Aid

Example of usage

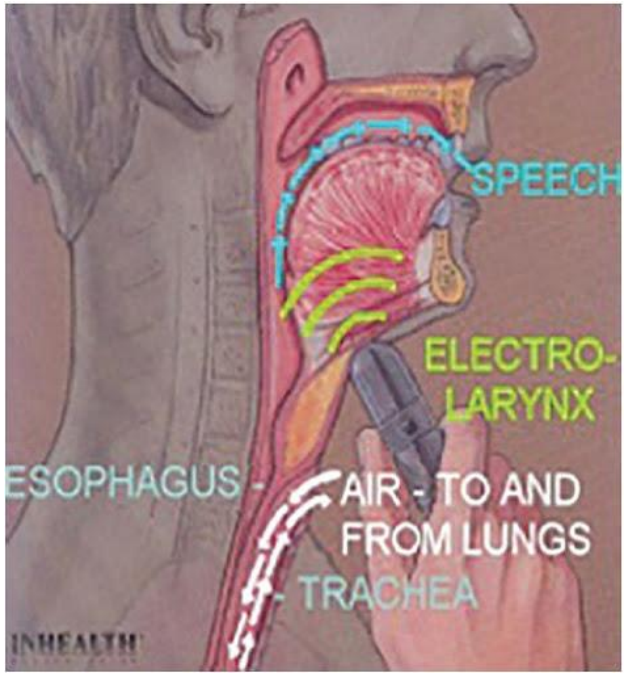
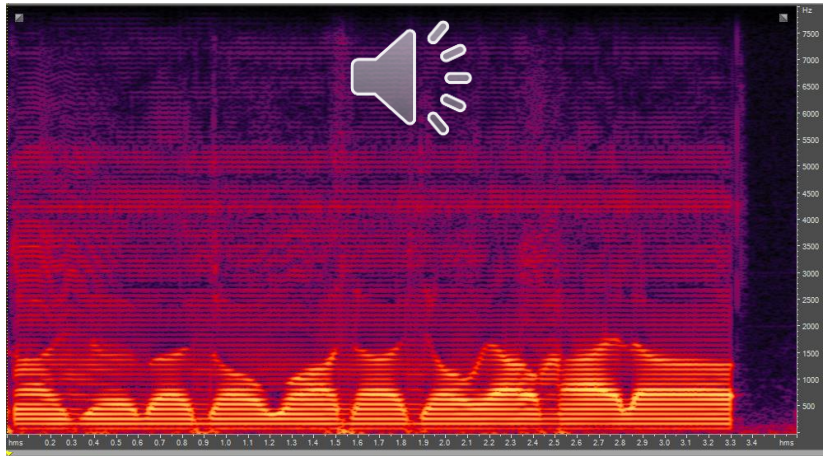


Image Source: https://www.brucemedical.com/nuvoisiii.html
http://mmj.eg.net/viewimage.asp?img=MenoufiaMedJ_2015_28_4_800_173591_u1.jpg

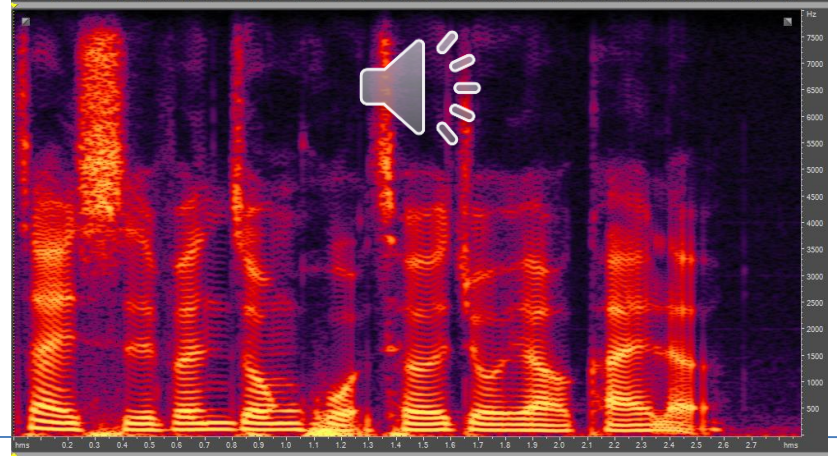
Throat EL speech example

Text: 再十分鐘火車就要開了

EL Speech 1

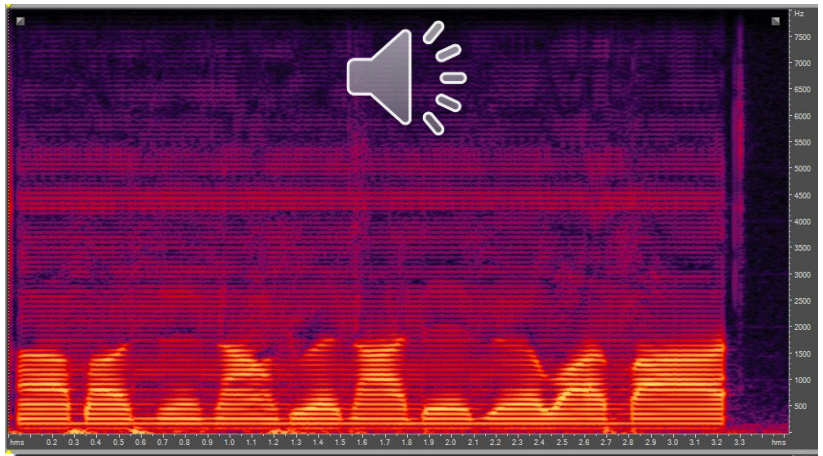


Speech 1

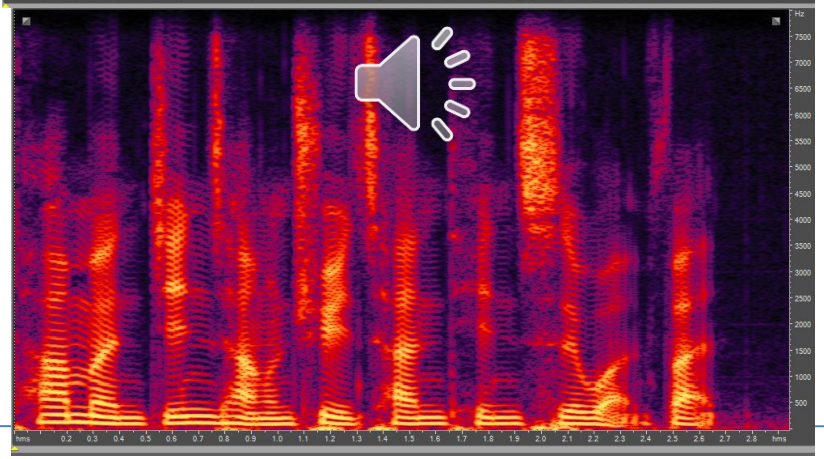


Text: 他們經常去餐廳吃晚飯

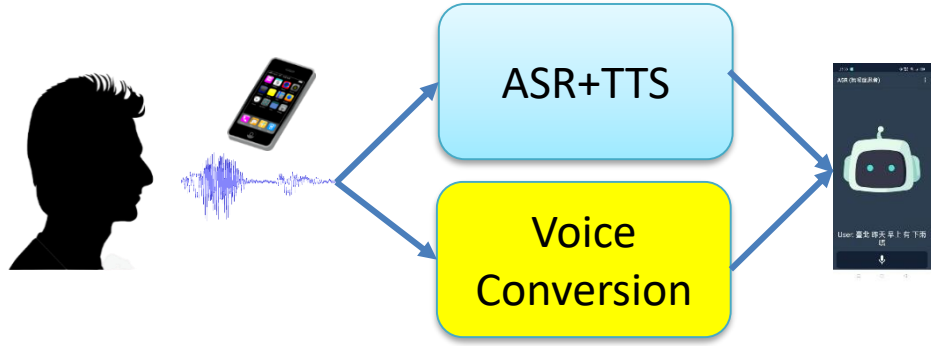
EL Speech 2

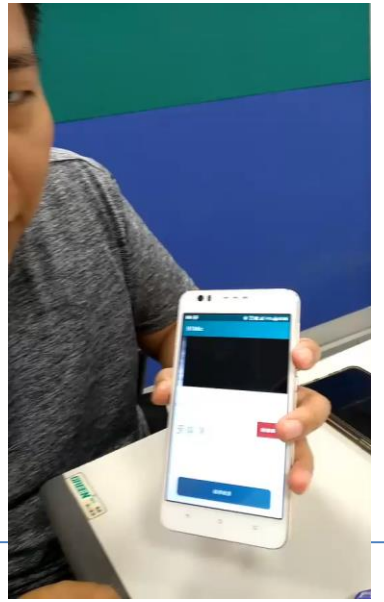


Speech 2



TSVoice: Electro-larynx Speech Recognition/Conversion
經鼻電子喉發聲輔助系統





漸凍人的語言溝通

「想做的都不能做、想說不能說，很嘔！」是許多漸凍人病友的共同心聲。

ALS 患者因為舌頭肌肉萎縮，舌頭動作遲緩，所以構音困難、說話不清楚。同時手和手指的肌肉也會無力，無法筆談，造成溝通困難。

漸凍人溝通新科技！　用「腦波」就能發聲

腦波溝通系統　簡易意念想像傳遞訊息

漸凍人語音系統　客製增溝通溫度

霍金21歲罹患ALS，醫師說他可能是世界上活最久(76歲)的漸凍人。