

Natural Intonation



September 2012 updating of *Luciano Canepari's*
canIPA Natural Tonetics

from Ch 12-14 of the book:
Natural Phonetics & Tonetics



12. Microstructures

(mostly: syllable, stress and tone)

Syllables

12.1. The SYLLABLE is a phonetic reality (both in an auditory and an articulatory sense), which is present in the linguistic consciousness of all people, no matter what language or dialect they speak, and whether they are literate or not. It is the fundamental unit of the spoken language – the smallest one capable of constituting an utterance by itself, such as *Yes* or *Here*, &c.

A syllable is made up of one or more phonic segments, which have a good deal of cohesion and coarticulation among one another. Syllables are also the groups into which we instinctively separate words, when we are speaking on the phone and the line is very bad (because of problems with static or interference).

For purposes of scientific analysis and description, it is normal to consider single segments (ie phones) as the minimal units of phonetics, such as *m*, *p*, *b*. And it is even possible to work on the level of components (or *phonetic features*), such as *bilabial closure* for *m*, *p*, *b*. However, actual speaking is carried out through full syllables (even if the syllables are made up of only one –short– phone, as in the Italian words *e*, *o*, *a* [e, o, a] ‘and, or, to/at’), whether these are stressed or not. The smallest isolated forms in English can be words such as *a*, *I*, *owe*, *awe*, *ah*, which are diphthongs or long vowels [‘eɪ, ‘aə, ‘ɜʊ, ‘ɔː, ‘ɑː], but in connected speech *a* is [ə].

Syllables make up in turn part of larger groups, constituted by sequences of syllables linked together by phenomena of assimilation or coarticulation: RHYTHM GROUPS (or *stress groups*).

In a rhythm group, one syllable has greater PROMINENCE with respect to near-by ones. The number of syllables making up a rhythm group can vary (depending on the speed of speaking and the particular message, as well as the language in question) from one syllable to around ten, with an average of 3-6 syllables per rhythm group.

12.2. The syllable can be considered as the result of the coordinated movements of the respiratory, phonatory, and articulatory mechanisms. Increases in general muscular tension, in expiratory pressure, in phonatory energy, and in the opening of the articulatory organs converge together.

These increases are followed immediately by corresponding reductions in the various areas. This is the point of view of *production* (ie the speaker); while from the point of view of *perception* (ie the listener), everything becomes transformed into sound waves, which travel to the auditory mechanism of the listener.

It is important that we clear up, once and for all, that phonetically the concept of a ‘syllable’ is rather different from that of grammar (not to mention poetic meter)! It is therefore essential to distinguish between syllables in a traditional sense

(*graphic* syllables) and natural *phonic* syllables. For this reason, it is entirely useful to speak of *phono-syllables* and *grapho-syllables*. In Italian, we can exemplify the difference with the word *festa*, with syllable division /fɛs-ta/ (phonic) and *fe-sta* (graphic). It ought to be unnecessary to mention the absurdity of the grammatical syllable division, given that the pronunciation is unmistakably [fɛsːta].

However, it is very difficult to defeat noxious scholastic ‘beliefs’, since they are inculcated at an early age. Grammarians and poets have, for centuries, convinced people that words like Italian *mai* rightly have only one syllable, whereas others, like *mia*, have two! Scientifically, there is nothing which is farther from the truth! In fact, we have: *mai* /ˈmai/ [ˈmaːi] and *mia* /ˈmia/ [ˈmiːa] (at the end of a line of verse, as well).

It is so simple and natural. Yet, in schools the official ‘story’ is that *mia*, *mie*, *mio* have two syllables (because they are made up of *mi-* and *-a*, *-e*, *-o* respectively), whereas the more ‘solid’ *miei* has only one, because it does not alternate with other forms! In reality, there is only one PHONO-SYLLABLE in all of these cases: /ˈmia, ˈmie, ˈmio, ˈmjei/ [ˈmiːa, ˈmiːɛ, ˈmiːo, ˈmjɛːi].

Scale of syllabicity

12.3. Within each phono-syllable, there is a considerable degree of correspondence between the openness or closure of the PHONOARTICULATORY ORGANS (ie *phonic production*) and the scale of syllabicity of the different phones (ie *auditory perception* of the sound wave). In fact, under *equal* conditions of stress, length, and pitch, the more open and more voiced phones are more *perceptible* (in other words, they are audible at a greater distance, ^{or} more distinctly). Clearly, in order to test this condition, a notable distance should be considered. In fact, if the distance involved were too close, it would be possible to get the impression that [ʃ] is more ‘sonorous’ than [a], particularly thinking of the ‘example’ of *shh!*, which is capable of getting a whole room full of people talking in groups to be quiet.

However, listening carefully (and if we look, also carefully, at a good transcription), we can see that *shh!* normally corresponds to [ʃʃʃʃ], in short something quite a bit different from just [ʃ]. This last would be plain, ‘non-syllabic’, voiceless, and short, without any particular stress –even wishing to transcribe it as [ʃ]– and in any case, without pitch, since the vocal folds do not vibrate. Here, however, a simple –‘egalitarian’– [a] is definitely much more perceptible (especially at a distance of 30-50 feet).

At this point, we need to make a brief terminological and conceptual digression, because, as could have been expected, the worst term imaginable (namely, ‘scale of sonority’) is also the most widely used, in the scientific literature as well. It is true that we are speaking of more or less ‘sonorous’ *sound waves*, but this in the very generic sense of phonic emissions, or in other words, actual utterances. Utterances are naturally composed of voiced phones, but also of voiceless ones, so there are a few problems here.

It would be a step in the right direction to get rid of the other expression (which

for a while seemed to us to be reasonably appropriate), namely ‘scale of perceptibility’. Although this term does not have the inevitable ambiguity inherent in ‘sonority’, it still runs the concrete risk of provoking the misunderstanding discussed above. In fact, it is fundamental to remember that absolutely *equal* conditions of stress, length, and pitch are required.

12.4. Therefore, it is more appropriate –and even necessary– to speak of the SYLLABILITY SCALE. Given the phones making up a particular utterance, the purpose of this *scale* is to make it possible to pick out the individual syllables. The division into syllables of an utterance is determined by the *syllable nuclei* (the maxima, the heights, the peaks, the apogees), in their contrast with *syllable boundaries* (the minima, the depths, the troughs, the perigees), where the actual division takes place.

With these considerations in mind, phones produced with greater articulatory (mouth and jaw) opening are more ‘perceptible’ (and consequently more apt to constitute the *nucleus* of a *syllable*) than ones with less opening. The same is true of phones with voicing (*ie* vibration of the vocal folds), with respect to voiceless ones (*ie* without this vibration). Along the continuum between voiced and voiceless phones, syllabicity diminishes constantly through the intermediate phases: *voiced, voiced lenis, mixed, voiceless lenis, voiceless*.

We therefore present now the syllabicity scale (going from the greatest to the least, cf fig 12.1). Every phonic syllable is, therefore, constituted by phones grouped together according to this scale, in such a way that the most perceptible ones (*ie* the syllabic NUCLEI) are in the center, while the less perceptible ones (*ie* the syllabic MARGINS) are on the boundaries.

fig 12 .1 Scale of syllabicity (with some examples).

1. <i>First group</i> : VOCALIC	
open vocoids:	[æ, a, ɒ; ε, ɐ ɔ; ã, ã̃, õ]
mid vocoids:	[ɛ, ɜ, σ; e, ə, o; ě, ẽ, õ]
close vocoids:	[ɪ, ɨ, ʊ; i, ɨ, u; ĭ, ĩ, ũ]
intense (<syllabic>) contoids:	[m̩, n̩, ŋ̩; ɹ̩, ʃ̩, l̩]
2. <i>Second group</i> : SONANTIC	
median or lateralized approximants:	[j, ɥ, ɰ, w; ɹ, ɻ]
trills, taps, flaps:	[r, R; ɾ, ɽ; ɹ, ɻ]
laterals:	[l, ʎ, ʟ, ł; λ, ʟ]
nasals:	[m, ŋ, n, ɲ, ɳ, ŋ]
3. <i>Third group</i> : CONSONANTIC	
peripheral approximant:	[β, ɸ, δ; ɣ, ɣ̃, ɦ]
constrictives:	[v, ð, z, ʒ, ʃ, ʂ]
stop-strictives:	[dʒ, dʒ̣, dʒ̥, dʒ̧, gʃ, gʃ̧]
stops	[b, d, ḍ, d̥, ʈ, ʈ̣]
4. <i>Fourth group</i> : NON-SONANTIC	
vocoids:	[i̥, u̥, ʊ̥, ɪ̥; e̥, ɛ̥; ɔ̥, ḁ]
sonants (or sonorants):	[r̥, R̥, ɹ̥; l̥; ɦ̥, ɰ̥, ɻ̥]
continuant (approximants/constrictives):	[f̥, ʃ̥, ʂ̥; s̥, ʂ̥, x̥]
momentary (stops & stop-strictives):	[p̥f̥, ts̥, tʃ̥, kx̥; p, t, ʈ, k]

In most cases, there are no problems, and the syllables form units with perfect internal coherence around the nucleus, moving from the most marginal phones (in both directions): *blank* ['blæŋk], *cleft* ['khlɛft], *ground* [ˈɡɹɑːʊnd, 'ɡɹ-], *final* ['faɪnəl], *written* [ˈɹɪt-ɪn, 'ɹ-]; Italian: *quando* ['kwɑnɔːdo], *tronfio* ['trɔmːfjo]; German: *Esel* [ˈʔɛːzl], *haben* ['haːbm]; Czech: *Petr* [ˈpɛːtʃ], *krk* [ˈkrk], *vlk* [ˈvɛk].

12.5. In all languages, however, there are cases –occurring more or less frequently– in which this order is not fully respected: *splash* ['splæʃ], *six* ['sɪks], *mere* ['mɪvə]; Italian: *sta* [ˈsta], *sberla* [zˈbɛrːla], *mia* [ˈmiːa]; German: *Stadt* [ʃtɑt]; Russian: *Pëtr* (Пётр) [ˈpɔtʃ]. When /sC, zC/ are preceded by /V/, they have the natural syllabification, which we have already shown in other transcriptions: Italian *questa* [ˈkwɛsːta], *una sberla* [uːnɑzˈbɛrːla], unless (as in English) the lexical syllabic structure does not take precedence: *a splash* [əˈsplæʃ], *six eyes* ['sɪks 'aɪz].

From a phonic point of view, however, every part not divided by a hyphen (or by a space) is a syllable. In fact, even though [s, z, ʃ] are more perceptible than [p, b, t], and [æ, a, ɔ] are more so than [i, ɪ, u], &c, we have no division into distinct phonetic syllables, provided that the elements in question belong to the same subdivision, out of the four in the scale, *ie vocalic, sonantic, con-sonantic, and non-sonantic* (note Latin *consonare* ‘to sound together’, since *con-* = ‘with’). The preceding is true as long as there is no increase of stress present. In fact, in [ˈja, iˈa], the first sequence constitutes a single syllable, while the second sequence constitutes two different syllables. As we have seen, there is just one syllable in Italian *mai* /ˈmai/, *vuoi* /ˈvwoi/, *miei* /ˈmjei/, and even in *mia* /ˈmia/, or in British English *mere* /ˈmɪə/ [ˈmɪvə]. This is true because, even though [i, a] have different levels of openness ([i] is closer than [a], and less ‘perceptible’, but it functions as a nucleus), they are both part of the same group: the *vocalic*. However, in the Italian of Naples (in a tune), we have two or three syllables in these cases, due to stress differences: *mia* [ˈmɪˌiːa], *mai* [ˈmɑˌiːi], *vuoi* [vuˌoːiːi], *miei* [miˌɛːiːi].

12.6. Let us consider some examples from English: *few* [ˈfjuː], *Swiss* [ˈswɪs], *piano* [phiˈænɔ, ˈphjænɔ], *situation* [ˌsɪtʃuˈɛɪʃən], *cooperative* [kɔˈɒpɹətɪv], *co-op* [ˈkɔˌɒp]. In English as well: *splashed* [ˈsplæʃt] and *sixth* [ˈsɪksθ] form only one syllable, because /sp, ks/ belong to the same group: the *non-sonantic*; all of this even though [s] is relatively more perceptible than [p, k]. The same is true of [ˈkɔˌɒp].

In French, words such as *quatre, peuple* /ˈkatʁ, ˈpœpl/ are monosyllabic, because [χ, l] are voiceless just like [t, p]: [ˈkatχ, ˈpœpl]; they are therefore in the fourth group (*non-sonantic*). Moreover, *double* [ˈdubl] is monosyllabic too, because [l], being voiceless (and *non-sonantic*), is less perceptible than [b] (in the third group, *con-sonantic*). If voicing is preserved (as would be expected phonemically), we have, instead, bisyllables: [ˈkatʁɔ, ˈpœplɔ, ˈdublɔ] (native speakers can even utter [ˈkatʁ, ˈpœpl, ˈdubl], still as two syllables, especially in protunes), as in Russian too, for the variant *Пётр* [ˈpɔtʃ], even though here the voicing is only partial.

In Japanese, voiceless (lenis) vocoids are still capable of constituting the syllabic nucleus, because the nearby contoids are completely voiceless and consonantic (and are therefore less perceptible, even while being part of the same group, *ie non-*

-sonantic): *kite* [ci₂t₁ɛ] /kite/ (or /kite/, of a different meaning, cf § 12.2.6.2 in *HPr*). The Japanese auxiliary form *desu* /de.su/ is bisyllabic until it remains [d₁ɛ.su₂], but it generally becomes monosyllabic [-d₁ɛs₁, -d₁ɛs].

Therefore, even while whispering (ie with voiceless lenis phonation), syllables remain intact, as is well-known. In fact, even if the message as a whole is less perceptible than one in a normal voice, the same differences in the scale mentioned above remain valid. When pronounced, a voiceless lenis, a vocoid, or an intense contoid, is still more audible than any other (normal or 'non-syllabic') voiceless contoid, whether lenis or not. The reader should try whispering, for example: *Yes, you're right* – *Oui, t'as raison* – *Sí, por supuesto* – *Sí, esatto*.

Syllabication (or syllabification): division into phono-syllables

12.7. In different languages, phonic syllables take more or less different forms, because they depend upon the phonological systems present, and on the languages' phonotactics (ie which combinations are normal or possible). Differences involve the number, order, and type of segments allowed, as well as the way in which syllabication is accomplished.

Within considerations of general phonetics, there are certain possibilities, which are more frequent and 'normal', and therefore defined as UNMARKED. These facts should be omitted from transcriptions, except in order to show intentionally the differences between an unmarked syllabification and other ones. It is, instead, important to note cases of MARKED syllabifications: those which are less frequent (or 'normal') in the languages of the world. In ordinary transcriptions, even these syllabifications can be possibly omitted, if they have been adequately explained. However, it should be emphasized that correct syllable division is very important for describing and pronouncing a given language well. Sometimes, it is neither easy nor simple to determine and perceive where the boundary between two phonic syllables is located, because of coarticulation, too, not only in unstressed syllables.

As we have mentioned, syllabic nuclei contain a notable amount of internal cohesion, which can also include coarticulatory transitions between the boundary elements of the single nuclei. The reader should remember that there is no real break between one syllable and the next (whether the boundary is marked with a hyphen or not), but a mere lessening of energy. It is important to carefully consider cases with a *simple* contoid (preceded by a stressed vocoid which remains phonetically short, and followed by vocoids, or by [central, lateral, trill, or nasal] approximants), or with a *geminate* or a *lengthened* contoid. In this last situation, there are respectable differences between Italian, Swedish, Finnish, and Japanese. A final point is that the hyphen is usually not used when there are prosodic symbols as well (especially for stress or pitch) in the same point, because in such circumstances it would be redundant and awkward.

12.8. Let us now consider some examples: *pepper* [p₁hɛp₁-ɐ], *paper* [p₁hɛɪ-p₂ɐ], *coble* [k₁hɜɔ-b₁ɫ, k₁hɔb₁-ɫ], *unyoke* [ɛp₁'jɜk], *anew* [ə'njɹu], *penknife* [p₁hɛn₁.næf],

bookcase [ˈbʊkˌkeɪs]; Italian: *pepe* [ˈpeːpe], *farò* [faˈrɔ], *acre* [ˈaːkre], *conio* [ˈkɔːnjo], French: *chapeau* [ʃaˈpø], *passer* [paˈse], *étoile* [eˈtwɑl], *noblesse* [nɔˈblɛs], *panier* [paˈnje]; Arabic: *atraab* [atˈrɑːb], *tanyiil* [tanˈjiːl]; Hindi: *vaakya* [ˈβaakˌjɑ], *aadnaa* [ˈaadˌna]. Moreover, Italian *penna* [ˈpenːna], *motto* [ˈmɔːto]; Swedish *penna* [ˈpʰenˌna]; Norwegian: *atten* [ˌatˌtɛn]; Finnish *tässä* [ˈtæʃːæ]; in Japanese, we have *motto* [ˈmɔːtɔ] (§ 12.3.2.1-4 of HPr).

Generally, those contoids, if any, which precede the nucleus of a given syllable have a more energetic articulation than those which follow it. In many languages, including most Germanic ones, stop and stopstrictive contoids occurring at the beginning of stressed syllables tend to be ‘aspirated’: *repeat* [ɹˈpʰiːt], *club* [ˈkʰlʌb], *chin* [tʃʰɪn]; or German *Betrieb* [bɛˈtʰɛːp], *kaum* [ˈkʰaʊm], *Pferd* [ˈpʰfɛːɐ̯t].

12.9. It is advisable to use rigorous terminology when referring to the structure of different kinds of syllables.

Thus, a SIMPLE (or mononuclear) syllable has a single vocalic element, a COMPOUND syllable, on the other hand, has a vocalic element which is (phonemically) long, or doubled (or geminated), or else a (true) diphthong (or triphthong): *cry*, *crying*, *layer* [ˈkʰɹɪə, ˈkʰɹɪəɪŋ, ˈleɪə] ([ˈkʰɹɪə, ˈkʰɹɪəɪŋ, ˈleɪɪ] American English); Italian: *avere* [aˈvɛːre]; in neutral Italian: *mai*, *mia* /ˈmai, ˈmia/ [ˈmaːi, ˈmiːa]; in some regional Italian pronunciations: *no*, *mai*, *mia* /ˈnɔ, ˈmai, ˈmia/ [ˈnɔɔ, ˈmaai, ˈmiia]; in German: *Eis*, *treu*, *treuer* [ʔaes, ˈtʰɛɔʏ, ˈtʰɛɔʏɐ̯] or [ˈtʰɛɔʏʌ].

An intense contoid, such as [ɹ], or [ɹ̥], can be considered appropriate as elements making up parts of diphthongs and triphthongs, in the real phonic sense of the terms (cf § 5.2-3), while we do not have ‘diphthongs’ in: *yes*, *wit* [ˈjɛs, ˈwɪt]; Italian: *ieri*, *può* [jɛːri, ˈpwo], nor ‘triphthongs’ in: *cube* [ˈkʰjuːb], or in Italian: *miei*, *vuoi* /ˈmjei, ˈvwoi/ [ˈmjɛːi, ˈvwɔːi]. In fact, [j, w] are actually plain (‘non-syllabic’) contoids.

Moreover, a CHECKED (‘closed’) syllable ends in one or more consonantal elements, as in: *texts*, *glimpsed* [ˈtɛksts, ˈɡlɪmpst]; Swedish: *skälmskt* [ˈʃɛlmskt, ˈʃ-, -st]; Italian: *cantan(te)*, *pas(ta)*, *trop(po)* [kanˈtanː(te), ˈpasː(ta), ˈtrɔpː(po)].

A FREE (‘open’, ‘unchecked’) syllable ends, instead, with a (simple or compound) vowel element: *I*, *me*, *you*, *potato* [ˈaː, ˈmri, ˈjmu, pʰəˈtʰeɪtɔ]; Italian: *so*, *sai*, *sono*, *causa* [sɔ, ˈsaːi, ˈsoːno, ˈkaːuza].

Finally, it can be useful (especially to determine the location of stress in certain languages, such as Arabic, Hindi, and Latin) to distinguish between LIGHT (free simple) syllables: *visi(bil)ity* [ˌvɪzə(ˈbɪl)əti], Italian *rivedere* [ˌriveˈdeːre], MIDDLE ones (free compound, or checked simple ones): *sorrow* [ˈsɔɹ-ɔ], *solo* [ˈsɔ-lɔ]; Italian *laurea* [ˈlaːrea], *cantante* [kanˈtanː(te)], or HEAVY ones (checked compound, or bi-checked/tri-checked ones, &c): *hands*, *(e)xempts* [ˈhændz, (ɪɡ)ˈzɛmpt, (wɪɡ)-] Italian *aus(pico)*, *ins(tallo)*, *frain(teso)* [ˈausː(piko), ɪns(ˈtalːlo), fraɪn(ˈteːzo)].

Syllables & the speech chain

12.10. There are cases of contoid sequences (in word-initial or word-final position), which belong to the same phonic syllable. However, if the same contoid se-

quences (or other similar ones) occur between vowels ^{or} intense contoids, then the sequences in question generally become split up into different syllables, according to the principles of nucleus and syllabic boundaries perceptibility. In sequences such as /Vsp, spV/, as in *spell*, *wasp*, or in Italian *sparo*, /s/ and /p/ belong to the same syllable (even though the stress mark is inserted between them in Italian, because we have *uno sparo* [u-nos-'pa:ro], cf § 3.3.1.3 of *HPr*): ['spɛt̪, 'wɒsp; s'pa:ro] (Ch 2 of *HPr*, on English, will explain our use of /IVt̪/, instead of the 'traditional' /IV/ [IVt̪]). However, in *hospital*, or in Italian *dispari*, the phonic syllable boundary is between the /s/ and the /p/: ['hɒs-pɪ-t̪; 'dis:pa-ri]. In addition, in an Italian phrase like *due strani film americani* 'two strange American films', we have [dues-'tra-ni 'fil-ma ,me-ri-'ka:ni].

The syllables of an utterance can contain variously long vocoid sequences. In fact, sequences of phones which all belong to the first group on the syllabicity scale (ie *vocalic*, 1-4, also including the more frequent intense contoids), are not uncommon: *you DO know who I am* [ju'dʒu:nɔ(h)uææm]; Italian: *i suoi autografi* [i,swɔiau'tɔ:grafi] (possibly: [-ɔjau-]), *lo direi a Eugenio* [lodɪ'rei aeu'dʒɛ:njo] (possibly: [-ɛ'jæu-]); French: *papa a à aller à Auteuil* [pa'pa AAalea'œtœj] (also [-A'A-, -A'A-]).

Phonetically, the English words given below are monosyllabic (whether pronounced with two vocoids and one intense contoid, or with three vocoids, or else with two, or with one only) – the American different variants are added in brackets: *towel* [tʰaʊt̪, tʰaʊt̪], *hour* ['aʊə, 'a'ə, 'a'a, 'a:, 'ɑ:] (Am. ['aʊ]), *fire* [fʰaɪə, fʰa'ɪ, f'a'a, f'a:] (Am. [f'aɪ]), *lower* [lɔə, lɔ'ə, lɔ'ɜ] (Am. [lɔ]), *employer* [ɪm'plɔə, -ɔə] (Am. [-ɔə]).

12.11. In languages such as Italian, Spanish, and French, the division of an utterance into syllables takes no account of word boundaries. Therefore, the initial and final contoids of certain words can become part of different, but contiguous phonic syllables. In Italian: *un'altra* [u'nal-tra], *non è vero* [no-nɛv'vɛ:ro], *per andare a Roma* [pe-ran'da-re ar'ro:ma]; in French: *mes amis ont été là* [me-za'mi zɔ'te-te'lA, -za'mið-te]; in Spanish: *los hombres iban al hospital* [lo'sɔmbre 'si'βa na'lospi'tal] (with [s] in American Spanish).

Other languages, such as English and German, preserve syllable boundaries, at least to some extent. This occurs especially in the case of stressed syllables. Sometimes it is possible to make semantic distinctions in this way (with varying degrees of consistency), as in these examples: *an aim* [ən'eɪm], *a name* [ə'neɪm]; *night-rate* ['naɪt-rɪt̪, 'naɪt-], *nitrate* ['naɪt-rɪt̪]; *I can see the meat* [ækn'sɪri ðə'mi:t̪], *I can see them eat* [ækn'si:ðəm 'i:t̪].

The different languages of the world not only have their own 'sounds' (and intonations), but they can also have different phono-syllable structures. For example, genuine Italian words (ie excluding recent or lofty borrowings, &c), do not contain sequences of differing CC which do not include /n, r, l, s/. This is because the (Italian) phonological system has assimilated the other sequences, producing geminates: *settantotto* [settan'tɔ:t̪o] (from se[pt]em and o[kt]o).

The Germanic languages, instead, can have a great number of heterogeneous groups of CC (and CCC as well), especially in Swedish and Icelandic. Slavic lan-

guages, such as Russian, also have a vast collection of consonant groups, with their own peculiarities. In Europe, sequences such as /ji, wu/ are generally unacceptable – in English they occur in a very small set of words, such as: *yeast*, *yip* [ˈjɪɪst, ˈjɪp], *womb*, *wolf* [ˈwʊm, ˈwɒlf]. But in Chinese, for example, /ji, wu/ is normal, for instance: *yī*, *wū* [ˈji, ˈwu].

12.12. We have mentioned (§ 12.1) that the syllable is the fundamental unit in spoken language. Therefore, in phonetic transcriptions the smallest isolated sequence occurring is, in fact, the syllable. This means that when plain ‘non-syllabic’ contoids occur, representing particular allomorphs, they must be joined to nearby vocoids. Cases include the article *l* in Italian and French, or cogemination in Italian (as in *a cena* [atʃˈtʃeːna], cf. § 3 of *HPr*). Contoids at the beginning (or end of words), which become part of different syllables, should also be joined in this way.

In normal speech, Italians do not distinguish between *Lavena*, *l’avena*, *la vena*. All of them are pronounced [laˈveːna] (the only possible difference is that the first form has a variant in /ɛ/ [ɛː]), and therefore the most ‘rigorous’ transcription would be to use the transcription above, in all cases. Possibly, in order to provide help for the learner, it would be possible to use a slur (symbol): [laˈveːna, l_aˈveːna, la_vˈveːna]. However, it is advisable to limit such expedients (which are quite a bit less useful than might seem) to a sort of explanatory experiment at the beginning of the learning process. Let us look at a similar case in French as well: *lavoir*, *l’avoir*, *la voir* [laˈvwaːʀ] ([laˈvwaːʀ, l_aˈvwaːʀ, la_vˈvwaːʀ]).

Length

12.13. The reader can compare § 6.4.1, where the basic ideas were first explained. The length in time of the articulation of a phone is measurable in hundredths of a second (ie centiseconds – $\frac{1}{100}$), or thousandths (ie milliseconds – $\frac{1}{1000}$) but it is sufficient –and more useful– to give relative indications, with respect to an average of $9\frac{1}{100}$ per phone, with oscillations ranging from about $6\text{--}12\frac{1}{100}$. This is, in fact, the necessary length required for adequate perceptibility of *short* phones, those considered to be of ‘normal’ length (except for taps, which –by their nature– cannot be longer than $3\text{--}6\frac{1}{100}$ without becoming trills or stops). Short phones have no associated diacritics, and are represented by the simple phonetic symbol [n], since they are just normal. If normal phones have less duration than the short degree, they are defined to be shortened and denoted by giving smaller versions of the original symbols as superscripts: [ṇ]. When, instead, the length is greater than the normal amount, we have *half-long*, *long*, and *extra-long* phones: [n̄, nː, nːː].

These LENGTH diacritics are called *semichrones*, *chrones*, and *superchrones*, respectively. On a phonemic, distinctive level, we speak of *chronemes*, as in German: *Stadt* [ʃt̄at̄], *Staat* [ʃt̄aːt̄]. We have already seen that in (British) English, forms like the following ones were transcribed traditionally: *ship* [ʃɪp̄], *sheep* [ʃiːp̄]; *look* [lʊk̄], *Luke* [lʊːk̄]; *not* [nɒt̄], *nought* [nɔːt̄]; *hat* [hæt̄], *heart* [hɑːt̄]. Since

then, transcriptions improved somewhat, giving: ‘/nɒt, nɔ:t/, /hæt, hæ:t/’; then another step forward was made, reaching the level of: ‘/ʃɪp, ʃi:p/, /lʊk, lu:k/, /nɒt, nɔ:t/’. However, we will never get tired of repeating that English ‘/i:, u:/’ are diphthongs (and not long monophthongs), and therefore a transcription like [ʃi:p, ‘lʊk] is much more appropriate, together with [ʃɪp, ‘lɒk; ‘nɒʃ, ‘nɔ:ʃ; ‘hæt, ‘hæʃ]. In fact, they should have the stress mark as well, even though they are monosyllabic, since other (monosyllabic) words are not stressed in sentence context (as we have said). It is, consequently, important to distinguish forms like *in* [ɪn] (prep.) from ones like *inn* and also *in* (adv.), both [ɪn:].

12.14. Distinctive length can also be manifested as gemination, especially for consonants, but also for vowels (in which case, there may even be a diphthong, whether mono-timbric or di-timbric, as has been seen in the case of English [ɪi, μu]): *part-time* [‘phɑ:tʃhɑ:əm, ‘phɑ:tʃhæm], *part-time job* [‘phɑ:tʃhɑ:əm ‘dʒɒb], *that time* [ðætʃhɑ:əm] (in modern Germanic languages, consonant gemination occurs only in lexical composition, or in sentence context): Italian: *cade* [‘ka:de], *cadde* [‘kad:de]; *ero* [‘ɛ:ro], *erro* [‘ɛ:rɔ]; Japanese: *gaka* [gɛ’kɛ], *gakka* [gɛk’kɛ]; *to-ki* [‘tɔ:cɪ], *tookki* [tɔ:cɪ]; Tamil: *paḍu* [‘paḍu], *paadu* [‘paadu]; *kanam* [‘kanam], *kannam* [‘kannam]; Hungarian: *vár* [‘var], *var* [‘vɒr], *varr* [‘vɒrr]; Russian: *žat’* (жѣтѣ) [ʒɛtʃ], *sžat’* (сѣжѣтѣ) [sʒɛtʃ].

A large number of languages use the length of particular segments distinctively, within their particular phonological system. In Arabic, Hausa, Hindi, Tamil, Hungarian, Finnish, and Japanese, for example, both consonant and vowel length are distinctive. Other languages, among which Italian, Bengali, Punjabi, and Amharic, use only consonant length for purposes of distinction. Still other languages, among which generally the Germanic ones, have phonemic opposition between short and long (or diphthongized) vowels. There are also languages such as Spanish, Romanian, modern Greek and Hebrew, Indonesian, and (Mandarin) Chinese, which do not use length distinctively, even though they can naturally have segments which are pronounced with a certain length for phonetic, phonostylistic (ie expression), or paraphonic (ie emotion or health) reasons.

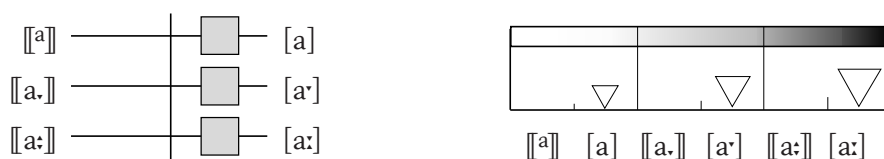
Two levels of (vowel or consonant) distinctive length are more than sufficient, in the languages of the world. In fact, what are supposed to be three distinctive levels of length in Estonian and Lapp are simply combinations of the two basic types. Examples from Estonian include: *jama* [‘jama], *jaama* [‘ja:ma], *jaama* [‘ja:ma], *kabi* [‘kabi], *kapi* [‘ka-pi], *kappi* [‘kap-pi].

Particularly in those languages which use pitch distinctively, there is a frequent use of *morae* (singular *mora*), which are the minimal units of length (Greek *μόρα* ‘division’). These units correspond to a short, or light syllable (such as /a/). A mid, or medium, syllable has two morae (made up of two vocalic elements, or of a long vocalic element, or of a short vocalic element followed by a consonantal one, which in turn may be normal –voiced or voiceless– or ‘syllabic’ (intense), as for example /aa, a:, an, aŋ, ad, at/). A long, or heavy, syllable has two vocalic elements and one consonantal one, or instead one vocalic and two consonantal ones (/aan, a:n, aaŋ, a:ŋ, ann, and, ant/).

In the systematic description of a language (and, better yet, while comparing multiple languages or regional accents), it could be useful to use a diagram which does a better job of showing the value of every symbol of length (every chronetic symbol). Therefore, beginning with the more specific symbols (already listed in § 11.21), we give the *chronogram* as well (in fig 12.2). For a level of (phonetic) length less than those of the three *fundamental* degrees, [a, a', a:], it is possible to use the three *supplementary* levels, [[a, a., a:]], respectively. We simply provide the diagram, with two possible types of scales to represent facts concerning length (their use can be easily intuited, even if they are not directly applied here).

In any case, we observe briefly that in Italian (respect to German, for example), it might be considered desirable to use [ː]: *bene, male* /bɛne, 'male/ [ːbɛ:ne, 'ma:lɛ] instead of [bɛ:ne, 'ma:lɛ]. Or in Spanish protunes, one could use [ː]: *quiero hablar* /'kjɛro a'βlar/ [ː'kjɛ.ro a'βlar], instead of [ˈkjɛ.ro a'βlar]. Finally, in American English (especially mediatic [treated in § 2.4.2.1-5 & § 2.5.2.4 of *HPr*]), it could be considered desirable to transcribe *beat, short* as [ːhɪɪt, 'hɪɪt; ːσ.ɹɪt], instead of [hɪɪt, 'σ.ɹɪt] (in neutral American pronunciation, we have: [hɪɪt, 'σ.ɹɪt]).

fig 12.2. Chronograms to show different degrees of length.



Pitch & tones

12.15. The melodic height (and movement) on a syllable is called TONE (cf § 6.4.4). Pitch is primarily determined by tension and vibration of the vocal folds and of the whole larynx. The more the folds are *tightened* or *loosened*, the more pitch becomes higher or lower. The effect can be accentuated by increasing the pressure of expiratory air and by raising the larynx; on the other hand, it can be diminished by reducing the pressure and lowering the larynx.

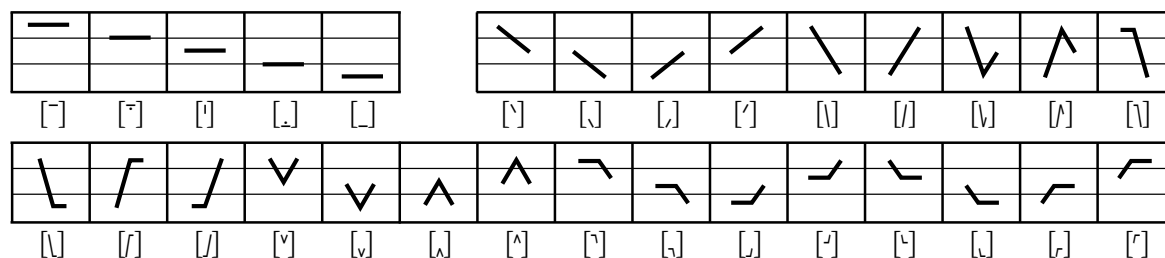
Pitch range is individual and relative, not absolute – to see this, it is enough to consider the notable differences between the average voices of men (low), women (mid), and children (high). Moreover, within each of these classes, there are innumerable possibilities for variation, including juxtapositions and exchanges from one class to another. See fig 6.17-18 and the tonograms from the phonosyntheses of tone languages (ie languages with ton[em]es), as well as fig 11.10-20 & fig 12.3-6 in *HPr*, for Mandarin Chinese and Japanese.

The range of pitch can be usefully divided into three juxtaposed bands: *high*, *mid*, and *low*. Of these three, the two bands on the ends are MARKED, while the central band is 'normal', neutral, UNMARKED. Thus in the mid band, no particular symbols for tones are required, and so [ˆ\$, '\$, ·\$, °\$] indicate syllables with mid tone and, respectively, *extra-strong*, *strong*, *medium*, *weak*, and *reduced* stress.

level tonemes). In other languages, such as Swedish, Norwegian, Croatian, Serbian, and Japanese, various combinations of pitch and stress on different syllables of the word, or rhythm group, determine the marked forms in comparison with the unmarked ones.

Normally, two different tonemes on elements which form a single syllable (with /V:, VV, VN/) assimilate one another. This occurs, for example, in Japanese: *kondo* [kɔ̃n.dɔ̃] (//. .//).

fig 12.3. Principal tones: symbols and tonetic characteristics.



Tonetic practice

12.17. In order to work with intonation (and tones), it is important to be acquainted with *pitch*, which is simply the height and melodic variation of speech (in one syllable at a time, stressed or unstressed). Pitch is determined by tension and vibration of the vocal folds and of the entire larynx.

The more the folds are tightened, the higher the pitch becomes; naturally, when they are loosened, the pitch becomes lower. The effect can be accentuated by increasing the pressure of expiratory air and raising the larynx. In the same way, the effect can be attenuated by lessening the pressure and lowering the larynx. At this point, it is absolutely indispensable for the reader to take a bit of time and investigate this mechanism.

It is necessary to become precisely aware of the correspondences between pitch movements and glottal tension and position (in the vocal folds and larynx). It goes without saying that the task can become appreciably easier if the reader records the attempts and listens to them calmly and repeatedly, possibly working with very small snatches of the recording at a time (and therefore pressing the pause button often).

In the beginning, it will be enough to make any sort of attempt whatsoever, to try to pass from one pitch to another, perhaps with big jumps to perceive the differences better. Afterwards, however, it will be necessary to work more systematically, attempting to execute more gradual and planned movements, after having accurately written them down.

The attempts should be carried out (whether planned ahead or not) by saying into a microphone what is meant to be recorded (or, more prudently, what has actually been recorded, since intentions are one thing, but actions are another). In this way, it is possible to compare the actual execution with what was planned.

These exercises should be organized so as to cover the full range of possibilities, from static pitch levels to various types of different and combined movements. Stress and length differences can be added as well.

At the beginning, it will be unquestionably necessary to plan on fairly long practice, with many repetitions. After covering the various combinations of pitch level and gradual movement which are possible, one should then move on to identifying one's own intonation characteristics.

12.18. It is clearly necessary to proceed while constantly recording oneself, and particularly to do so without becoming overly depressed. Discouragement often comes from hearing one's own voice and one's own pronunciation characteristics (which at the beginning can be a rather unpleasant experience), as well as from the feeling of not being able to do anything decently. We hear our voice on a cassette recorder –or, nowadays, from a computer sound file– in a different way than we are accustomed to hearing it. In fact, when we hear our own voice, we hear it not only from vibrations in the air, but also from vibrations which propagate through the bones of the skull, directly from the phonatory mechanism to the auditory one.

As for the other problem –our apparent perceptual and productive limitations– it is enough to remember that we are not at all used to efforts of this type, and therefore they require time and personal commitment. The tonograms of $\mathcal{C} 6$ & $\mathcal{C} 13$ (fig 6.17-21 & fig 13.1-9) will certainly be of help, as will the phonosyntheses of $\mathcal{C} 16-23$, or those in *HPr*; they are not at all superfluous, nor useless, nor inaccessible, either. It is sufficient to work at them without superficiality and without distrust – they will more than repay the effort!

The pitch range (that is, the full range from the highest pitch to the lowest pitch), therefore, is individual and relative, not absolute. As we have pointed out in § 12.17, there are noteworthy differences between the voices of *men* (generally low), of *women* (higher), and of *children* (still higher, cf fig 13.1), and there is a great range of variation even with respect to these norms. Each person must put a bit of effort into discovering and analyzing his or her own pitch range, which is more limited in extent than the range in singing.

Working as always with a computer sound program –or a tape recorder– the reader should be recorded while speaking 'spontaneously'. That is, one should not just read a written text chosen at random, but pronounce various words and sentences of different types, uttering them, however, as if they were occurring in natural conversation. For this purpose, the examples from fig 13.2.1 would work very well (although they represent just neutral British English intonation); they will, then, need to be listened to repeatedly, both as a whole and using the pause button to segment them. Of course, for other accents and languages, it would be a very nice thing to be able to rely on the corresponding curves. Many such curves, on tonograms, are already available from our books. Others can be found in some pdf files freely downloadable from our web site. Many more are appearing either in our site or in our books.

It will, then, be important to focus on the various details, which as the exercises proceed, will come out more and more clearly and seem more and more obvious,

even if they went completely unnoticed at the beginning. It is necessary to learn to listen to every single characteristic and to every component and variation, independently from the others, simultaneously present. The characteristics are woven together in a sort of tangled web, which is however harmonious and even melodious.

In fact, the reader must succeed in hearing, not just the entirety and not just generalities, but also in listening to the single characteristics, perceiving as many details as possible. Something similar occurs while listening to a full orchestra: it is possible to hear the music as a whole, but it is also possible to learn to recognize –and savor– every individual instrument.

All in all, the same is true of all phonetic characteristics. The only difference is that tonetic characteristics are effectively more complex, and therefore require a greater commitment.

12.19. After the initial experiments for discovering one's own pitch range in speech, the reader should choose sentences in which all the vowels are the same (phonically as well – thus, in Italian for instance, it is important to avoid mixing together /e, ε/ and /o, ɔ/). Possible examples for Italian include: *Quindici bimbi simili di Rimini*, or *Per prevedere tre sere vere*, or *Sono molto forforoso*, or possibly *Mangia la patata salata*, or *Sara sarà andata armata da Catania a Malaga* (with [ˈmaɲːdʒa, kaˈtaɲja]).

These sentences will not necessarily be among the most normal or probable in conversation. However, the fact of having all of the vowels of the same quality in all the syllables is a great help. The reason is that in this manner, we avoid the risk of being misled by the different pitch intrinsic to the different vowels (even if acoustically different contours can influence vocoids to some extent).

Therefore, the reader can proceed at this point with less difficulties, modifying the intonation (cf § 6.4.5.1-4 & ¶ 13), moving from a conclusive sentence to an interrogative one, or else a suspensive or a continuative one. The reader should use recordings as always, and listening to the results with care, multiple times. It is necessary to pay attention to all the differences in pitch, while leaving aside other characteristics for the time being. In languages which differ from Italian because they have different vocalic timbres in stressed or unstressed syllables, it could be difficult to obtain suitable phrases and sentences. Thus the words could be changed in order that the same vowels are used instead of the original ones, as for instance in [ɑːvːɑːnlɑː ˈtʰɑː ˈkɑːts ɑːndwɑːnˈdɑːg], [ɪvːɪnlɪ ˈtʰrɪ ˈkʰɪts ɪndwɪnˈdrɪg], [əvːən-lə ˈtʰə ˈkʰəts ɛndwɛnˈdɛg] (*I've only seen two cats and one dog*), or it could be possible to simply hum each sentence.

In any case, using our *intonation* system (and perhaps sentences of our own language), we should be able to establish what our low(est) pitch is by considering the end of a conclusive tune, such as (though here examples are still from Italian – with the same vocoid) *da Malaga* 'from Malaga' [daˈmaːlagaː], as well as the high(est) pitch in *da Malaga?* 'from Malaga?' [ɔːdaˈmaːlagaː], which will probably be on the last syllable, or the second or third to last, depending upon whether we have an interrogative tune of the rising type, as in neutral Italian and British English [ˈˈˈ], or one of the rising-falling type, as in [ˈˈˈˈ], [ˈˈˈˈˈ], [ˈˈˈˈˈ] or [ˈˈˈˈˈˈ],

[· ¯ ··], [· \ ··], [· ' ··], &c (here we show the ideal four-syllable pattern). Then we should try to discover our mid pitch, by seeking a point between these approximate (and individual) extreme points.

It is very useful to add voluntary modifications, even if this makes the utterance more forced – the important point is to explore new things constantly. The experiments can involve single words, or even single syllables, with all the vowels, the same as before (but naturally, in the future, it will be necessary to move on to more natural words and sentences, with different kinds of vowels present at the same time).

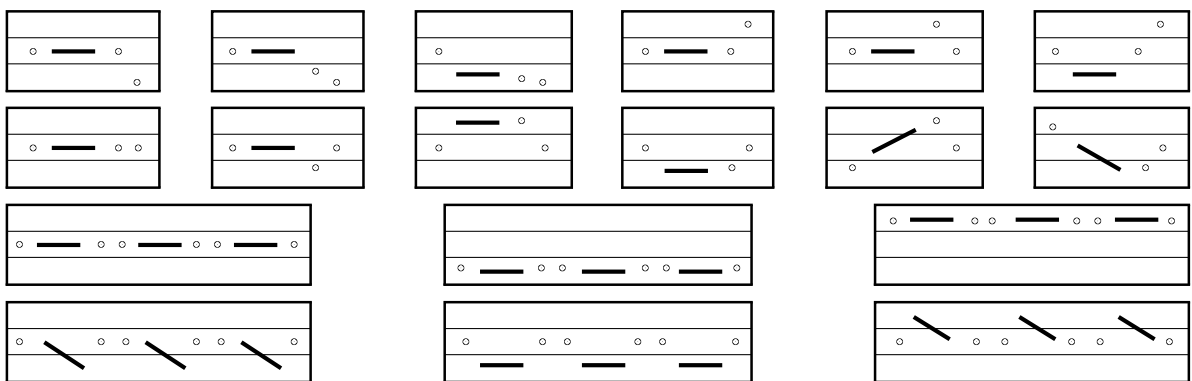
12.20. For *tones*, as well, it would be good to draw tonograms (the figures with the three juxtaposed bands, as in fig 12.5-7), and to mark what is heard when listening to the recordings.

Another possibility is to mark a pitch beforehand, and then work at trying to reproduce that exact pitch perfectly, listening to the recording immediately afterwards. Syllables like *la-la-la* can be used [lɑː lɑː lɑː], and then the pitch can be modified while maintaining the stresses, giving, for example: [_ _], [' '], [¯ ¯]; [_ '], [¯ '], [' ¯], &c.

More syllables will be added: [_ _ ¯ ¯], [_ _ ' '], and gliding tones will be added as well: [/ '], [\ '], [¯ \ '], [/ '], &c, including more complicated ones and compound ones: [\], [/], [\], [/], [\], [/], [\] &c (better than the older analytical signs [\], [/], [\], [/], [\], [/], &c, cf fig 12.5, choosing between the tunings and the tones – if present – of the phonosyntheses of *G* 16-23, or of *HPr*, and other books or pdf's of ours.

It is also possible to use nonsense words, such as: [phɑːʰɑːkɑː], [phiiʰiiki], [phɯʰɯkɯ] (similar to the patterns [pɑːtɑːkɑː], [piːtiːki], [puːtuːku] we have used in the audio recordings of *Phonetic Notation*, given in the bibliography).

fig 12.4. Tonograms for exercises with tones.



It will also be useful to perform exercises on *intonation* as well (cf fig 12.6), perhaps after having seen *G* 13, also with [phɑːʰɑːkɑː pɑːʰɑːkɑːʃɑː], [phiiʰiiki piiʰiiki-tʃii], [phɯʰɯkɯ pɯʰɯkɯtʃɯ], alternating between *tunes* like [· ' ·], [· ' ··], [· _ ·], and [· ' ·], [· ' ··], [· _ ·], and [· ' ·], [· ' ··], [· ¯ ·], [· ¯ ··], [· ' ·], [· \ ·], as a start.

And it would not be a bad idea to try *protunes* with various peculiarities, using [phɑːʰɑːkɑː pɑːʰɑːkɑː pɑːʰɑːkɑː], [phiiʰiiki piiʰiiki piiʰiiki], [phɯʰɯkɯ-

From a perceptual point of view, listeners normally manage to distinguish which syllables are stressed by putting together all the available elements of the utterance, in order to reconstruct and deduce the way in which they would themselves produce those elements to obtain analogous results. For practical purpose, we can speak of the process in simplified terms of *force* (production) and *volume* (perception). Therefore, once the prominence of a syllable has been established, excluding severe variations in pitch, length, and timbre (which are present, but occur in their normal, inevitable proportions), the relative differences with nearby syllables can be analyzed, in order to gradually modify the strength of the stress.

It is not rare to find people who cannot find the location of the stress in their own language, when vowel timbres have no significant differences between stressed and unstressed syllables. If they are asked to indicate which syllable is stressed in words like Italian *cavallo*, *margherita*, or even *fare*, *farai*, they may answer completely at random! They do not make mistakes, however, with *farà*, *così*! They are completely deaf concerning prominence (and also concerning the other things not shown in writing, in particular the timbres of *e*, *o*, *s*, *z* /*e*, *ε*; *o*, *ɔ*; *s*, *z*; *ts*, *dz*/). There is, however, one way (which goes back at least to certain Latin grammarians and teachers) to learn which syllable is stressed. The person is asked to pronounce the words in question as if shouting them to someone quite far away. Immediately, or immediately after the first try, it will be enough to *pretend* to shout.

Quickly enough, anyone can learn in this way how to tell correctly which syllable is stressed: [ka'vallo, margerita, fare, farai; fara, ko'zi]. In fact, in this way, the stressed phonetic syllable is notably lengthened, and it receives a generally high(er) prominence which is decidedly perceptible. This remains true even though the other syllables (particularly the last one) also gain in prominence – they still gain less prominence than the truly stressed one: [ka''val:lɔ, mar'ge''ri:ta].

Naturally, all this is true of one's own language and, in this case, for an Italian native speaker. However, if this speaker, for example, pronounces *ippodromo* or *gratuito* as [ɰippo'drɔ:mo, ɰgratu:'itɔ], then this is the stress location that will come out of the exercise (in this case, it is a pattern which should be decidedly avoided [as the present author has indicated in the *DⁱPI* – Dictionary of Italian Pronunciation – marking these realizations with 'ɰ']). For this reason, in order to get reliable answers, it is necessary to consult a good dictionary, or better still, a dictionary of pronunciation. This is true of native speakers as well, and therefore even more so of foreigners.

In this last case, the method just indicated for finding the stress in a word, will obviously not work. In fact, what would be determined, at most, is the position (correct or otherwise) that the *foreigners* attribute to the given word, according to their personal experiences and knowledge as non-natives.

12.22. In considering normal denotative utterances, it is important to pick out which syllables bear the SENTENCE STRESS (ie *ictus*). These are syllables which have *strong*, or *primary* stress: [!\$]. Other syllables have less strong stress: *mid*, *medium*, or *secondary*: [!\$], or *weak* ('without stress' and –rather– without a mark): [\$]. This last corresponds to the *neutral*, UNMARKED, level of force for a syllable. The sylla-

ble remains, however, fully audible, even in cases when the nucleus is *attenuated*, as with [ə, ŋ], &c. It is also possible to have syllables with *reduced* stress, written [°\$], that is, less than the normal weak stress; this is true particularly in rapid speech. On the other hand, there are also syllables with *extra-strong* stress, [!"\$], which is *emphatic*.

The different languages can use (the position of) stress in a word as a distinctive element, in a more or less extensive way. From a phonemic point of view, it is appropriate to try to understand the unmarked structure of every language with FREE STRESS (ie, not automatically linked to a particular syllable, for phonological reasons, or reasons of syllable length). For example: *photograph* [fʰɔʃəgɹɑf], *photography* [fə'fʰɔgɹəfi], *photographic* [fʰɔʃə'gɹæfɪk], (*an*) *increase* [ˈɪŋkɹɪs], (*I*) *increase* [ɪŋ'kɹɪs]; in Spanish: *término* [tɛr'mino], *termino* [tɛr'mi'no], *terminó* [tɛr'mi'nɔ]; in Italian: (*io*) *capito* [kɑ'pito], (*ho*) *capito* [kɑ'pito], *capitò* [kɑ'pito], *fotografo* [fo'tɔ:grɑfo], *fotografano* [fo'tɔ:grɑfɑno], *fotografare* [fo'tɔgrɑ'fɑre], *fotografò* [fo'tɔgrɑ'fɔ].

DISTINCTIVE (or *phonemic*) stress, as in some of the cases seen just now for English, Spanish, or Italian, could be called a *stresseme*, in opposition to mere *stress* (or instead, as will be seen very shortly, it would be possible to speak of the *dyneme*, in contrast to *dyne*, in order to use more careful and scientific terminology in this area as well).

At times, it is important to speak of *destressed* syllables, [°\$], not just unstressed ones (ie [°\$], but [°\$] in phonemic formulae). This refers to syllables which are normally stressed, but which become unstressed, or with reduced stress, in a phrase or sentence. This typically occurs because the words in question are weak monosyllables, or weak elements of lexical compounds (or else to lexemes which are weak in sentence context, for pragmasemantic reasons). Depending upon the languages, the words are typically grammemes and lexemes of little prosodic and pragmalinguistic importance, which have normal primary stress when pronounced on their own, in isolation, or as original forms. Languages can behave differently, in terms of both prosodic and articulatory reduction.

In Italian, for example, in cases like *è vero*, *tre volte*, *poiché*, *benché* we have full reduction from a prosodic point of view, but only partial reduction in terms of the articulations; in fact, we have [ɛv've:ro, trev'vɔl:te, pɔi'ke, bɛŋ'ke] (cf ¶ 3 of *HPr*). In Spanish, prosodic reduction of lexical monosyllables is less than in Italian (cf ¶ 6 of *HPr*), and the same is true of northern Italian dialects. In Catalan (and in some other Romance languages, including many northern and southern Italian dialects), only a restricted set of phonemes generally occur in unstressed syllables of bisyllabic and polysyllabic words, due to the presence of extensive neutralization, even to /ə/. Often, the vowel phonemes allowed are only /i, a, u, ə/ (Neapolitan), /i, a, u/ (Sicilian), or also /i, ʌ, u/ (neutral Catalan), or /ə, a/ (the Italian dialect of Bari, Apulia).

However, in these languages as well, destressed vowels are generally not a phenomenon having to do with phonic reductions of (isolated) words, but they rather follow the different phonic laws of the utterance (of the word in context and of *connected speech*), that is, of *phone groups*. In fact, although descriptions commonly emphasize the idea that neutral Catalan has only three V in unstressed syllables,

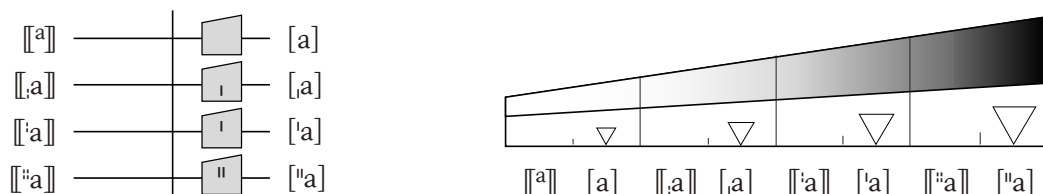
(/i, ʌ, u/ [i, ɤ, u]), in sentence context (and in the first elements in lexical compounds), we naturally find /e, ε, a, ɔ, o/ [e, ɛ, a, σ, o], as well, as in these Catalan examples: *és gros*, *com més*, *ha fet* [ez'ɣrɔs, kɔm'mes, a'fet].

12.23. fig 12.6 gives two diagrams (for two different types of scale) which could be useful for showing the peculiarities of stress in different languages, and thereby in making comparisons on general interlinguistic levels easier. In fact, the *dynogram* (or *stressgram*) makes it possible to observe the continuum of *dynetic* levels (that is, the force of stress, aside from simply functional *dynemic* peculiarities (of course, the word is derived from Greek *δύναμις* <*dýnamis*> 'power, force, strength') to which the normal markings can refer: weak stress, [], secondary (or mid) stress, [], primary (or strong) stress, ['], and emphatic (or extra-strong) stress, ["]).

In cases where intermediate gradations occur, they could be: half-mid stress, ['] half-strong, ['], and possibly, also half-extra-strong, ["]. An extra syllabic nucleus (usually epenthetic, with functions of support) can be called extra-weak. It can be generically marked with [*] (the same symbol as for perceptible explosion of stops, or other contoids, since the two phenomena coincide in production), or with a superscript of the vocalic symbol, a more articulatorily appropriate method (as we have done in Hindi, *HPr*, § 10.3.1.4).

Observing fig 12.6, we perceive the four *fundamental* categories: [a, ʌ, 'a, "a] (weak, mid, strong, and extra-strong), and the four *supplementary* ones: [ᵃ, ʌ, 'a, "a] (extra-weak, half-mid, half-strong, and half-extra-strong – added within ([]]). For each of these, the right half refers to the 'normal' position, with the cursor in the middle of it, while the left half is used for levels which are weaker than normal, should these occur. These supplementary intermediate gradations, [ᵃ, ʌ, 'a, "a] could be useful for explaining particular cases, as we have, for example, done with certain compounds in English (*HPr*, § 2.3.4.2), German (*HPr*, § 5.3.3.2 & 5.3.3.5), Spanish (*HPr*, § 6.3.2.2), Portuguese (*HPr*, § 7.3.2.2), and Hindi (*HPr*, § 10.3.2.5). The extra gradations should also be understood as effective possibilities for oscillation.

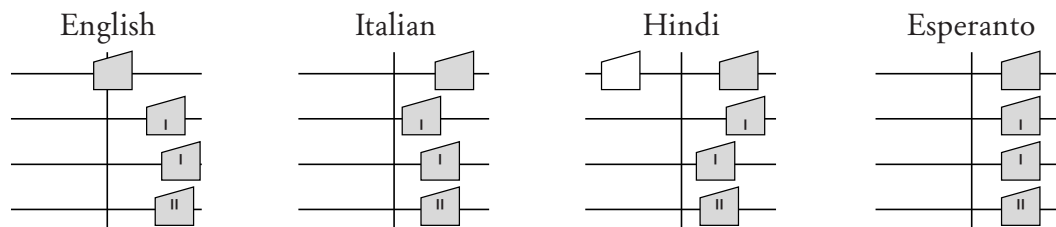
fig 12.6. Dynograms, for gradations in stress.



In a systematic treatment of the pronunciation of a language, such as in a specific monograph, it is possible to show greater detail. Just as much more space would be given to intonation than in the chapters of *HPr* (where the intonation is given 'embryonically', but with all of its potential for applications), it would be possible and useful to present the dynogram (and the chronogram) in their respective sections (or together with the tonograms). Here, we give in fig 12.7 a comparison between English, Italian, Hindi, and Esperanto. Several peculiarities can be noted,

including the extra-weak level in Hindi (indicated by the white cursor, used for very short vocoids serving only as support), and, in contrast to the others, the canonical regularity of Esperanto.

fig 12.7. Dynetic comparisons between English, Italian, Hindi, and Esperanto.



Stress (or word stress – mostly with English examples)

12.24. We know that (the position of) stress may be distinctive, in English: [ʊmˈpɔːt] *import* (noun, adj.), [ʊmˈphɔːt] *import* (verb); [ˈpreznt] *present* (noun, adj.), [preznt] *present* (verb).

English sentences usually keep the stresses of their words well, even in monosyllabic lexemes (ie lexical monosyllables), while monosyllabic grammes (ie grammatical monosyllables) lack any stress (as, in general, do unstressed syllables in polysyllabic words): [ˈsæmz bɔːt θɪri ˈnjuː smɔːl ˈkæt] *Sam has bought three new small black cats*; but we have: [ɪðəwəz ˈlɑːdʒ ˈkɪrənd ˈpɪpəl] *and there was a large crowd of people*, [bət ɪts ɪdʒ ˈgɪrənt ɪmˈpɔːt] *but it's of the greatest importance*.

However, in long words (and in sentences as well), many syllables with full vowels generally receive secondary stresses (especially when they occur near unstressed and reduced syllables): [ˌɔvəˈɛstɪˌmeɪt] *overestimate*, [vɪzəˈbɪləti] *visibility*, [ˌpɜːpənˌdɪkjələˈrɪti] *perpendicularity*, [ˌrɛkəˈnɪz] *recognize*, [ˈmɛksɪkə] *Mexico*.

In compounds, the more frequent structure is [ˈ\$, \$] (more rarely [ˌ\$, \$]: [ˌfɪfˈtɪn] *fifteen*). Sometimes, even [ˈ\$, \$] occurs, as in ‘collocations’ (or occasional –or free– compounds, which are, then, modifiable): [ˈblækˌbɜːd] *blackbird* (but [ˈblæk ˌbɜːd] *black bird*), [ˈbɒlˌdɒg] *bulldog*.

Of course, there are also many instances like: [ˌɪŋɡlɪʃˈtɪtʃə] *English teacher* ‘a teacher of English’ and [ˌɪŋɡlɪʃ ˈtɪtʃə] *English teacher* ‘a teacher who is English’.

Let us now consider compounds such as *first class* (noun and adverb) and *first-class* (adjective), and the collocation *first class*, in a sentence like *that was the first class to be considered*. From a phonetic point of view, they are alike: [ˈfɜːs(k) ˈkhlɑːs]; however, from a phonemic point of view, and for teaching and lexicographical purposes as well, it could be very useful to distinguish them as: /ˈfɜːstˈklɑːs/ (compounds: ‘/\$ \$’) and /fɜːst ˈklɑːs/ (collocation: ‘/\$ \$’).

Besides, patterns are flexibly structured. As a matter of fact, we have: [ˌfɪfˈtɪn] *fifteen* and [ˌpɛɪˈdʒ ˌfɪfˈtɪn] *page fifteen*, but [ˌfɪfˈtɪm ˌpɛɪˈdʒz] *fifteen pages*; [ˌbrændˈnjuː] *brandnew*, but [ˌɒˌbrændˈnjuː kəmˈpjuːtə] *a brandnew computer*.

Moreover: [ˌsekʊnˈdɪ] *secondhand*, but [ˌsekʊnˈhænd ˌkhlɜː(ð)z] *second-*

hand clothes; and [ðɛzɪˈstɪˌsɛkwɪnˈhæːndʒ] *they're all secondhand*; also [ˌɑːftəˈnɪˌmʊn] *afternoon* and [gɒdˌɑːftəˈnɪˌmʊn, gɒdɪ-, gʊdɪ-] *good afternoon*, but [ˌɑːftəˌnɪˌmʊn ˈθriː] *afternoon tea*.

A few cases can vary according to speech rate, but also whether they occur in tunes or protunes, as well as according to personal choices. Here, we will make use of different degrees of intermediate stress, too, which (without an emphatic one, [ˈ]) are, in descending order: [ˌ], [ˑ], [˒], [˓], [˔]. It is worthwhile observing nuances carefully: [dɛmənˈstɹɪɪʃn ˈɛksəˌsæzɪz, ˈdɛmənˌs-, ˈstɹɪɪʃnˈɛksəˌsæzɪz] *demonstration exercises*, [ˈɛləˌveɪtəɹ ˈɒpəˌreɪtɪv, ˈɛləˌveɪtəɹ ˈɒpəˌreɪtɪv, ˈɛləˌveɪtəɹɒpəˌreɪtɪv] *elevator operator*, [ˈlɑːtʃhaʊs ˈkhiːpɪv, ˈlɑːtʃhaʊsˈkhiːpɪv, ˈlɑːtʃhaʊsˌkhiːpɪv] *lighthouse keeper*; let us notice and compare: [ˈlɑːtʃ ˈhaʊsˌkhiːpɪv] *light housekeeper*.

12.25. To feel certain about the stress patterns of compounds, it is necessary to look them up in reliable dictionaries. But pronunciation dictionaries are not always the best choice, for this aspect, although, of course, they have to be consulted. We willingly recommend the Random House dictionaries which, for secondary stress, are almost perfect; of course, the stress patterns shown are American (and, practically, International) ones, but, in general, they may hold good even for British English, which, in the meanwhile, may have added ^{or} kept some other possible variants (mainly collocation-like, rather than compound-like, so less useful ones: *weekend, icecream, New York, New Zealand, New Hampshire...*).

In addition, the Oxford 'Advanced Learner's' dictionaries show the 'marked' cases of primary stress in several lexical collocations (which are quite unpredictable, syntactically, above all for foreigners).

In (dia)phonemic transcriptions such as ours, the most typical and numerous compounds are shown with a single primary stress /\$\$/; the secondary one is easily recoverable, because the second lexeme necessarily bears a secondary stress.

Vice versa, most dictionaries printed in the USA include secondary stress, /\$,\$/; but usually the non-IPA symbols they use put stresses *after* stressed syllables, unfortunately, not *before*, and simply through a difference in thickness (which, sometimes, is not evident enough, even with both of them *in praesentia*). As a matter of fact, we happen to find, eg 'in scrib' for our [ɪnˈskɹɪəb] *inscribe* and 'vizə bil'ə tē' for [vɪzəˈbɪlətɪ] *visibility*. But some American dictionaries are misleading, because they mark secondary stress for most unstressed syllables bearing full (unreduced) vowels.

Regrettably, mainly dictionaries published in the UK (even pronouncing dictionaries) do not use secondary stress wisely enough. As a matter of fact, a collocation like [ˈsɛntrɪˌtɪ ˈhiːtɪŋ] *central heating* is, usually, represented as */sentrəl ˈhiːtɪŋ/, exactly like *centralistic* */sentrəlɪstɪk/, which, of course, is [sɛntrɪˌtɪ ˈlɪstɪk].

However, the more they mark the better, even when things are predictable, provided they do so in an exact and accurate way. Indeed, didactic transcriptions (especially for beginners) should show several characteristics, with no absurd and groundless fear that they may confuse. In reality, too simple a transcription is less useful and, sometimes, misleading, too.

As regards diaphonemic transcriptions in compounds with suffixes, it is suffi-

cient to know which of them are always *non(half)stressable* ('/o\$/') and which are prosodically (*half*)*stressable* ('/o\$/'). As a matter of fact, the others, that have full vowels, are always (*half*)*stressable* ('/o\$/'). In addition to those with /ə, ɪ/, the following are always unstressed: [-ɪk, -ɪks, -ɪŋ, -ɪʃ, -ɪst, -ɪv, -ɪt] *-ic, -ics, -ing, -ish, -ist, -ive, -phil*: [ˈɪfɪəɪk, ˈphɒlətɪks, ˈlɪŋgʊərɪŋ, ˈjɛlɔʃ, ˈnɒvəlɪst, dəˈskɪpʃɪv, ˈæŋgləfɪt] *rhetoric, politics, lingering, yellowish, novelist, descriptive, anglophil* (for *-phile*, we have [-ˈfaɪl, -fɪt]).

Instead, the following are *half-stressed* (if preceded by an unstressed syllable), but *unstressed* (if preceded by a stressed syllable): [-hʊd, -ɪzəm, -aɪt, -aɪz, -ɪp, -juːt] *-hood, -ism, -ite, -ize (-ise), -ship, -ule*: [ˈwʊmənˌhʊd] *womanhood* (≠ [ˈtʃhaɪt(d)hʊd] *childhood*), [ˈtɪrəˌrɪzəm] *terrorism* (≠ [ˈbʊdɪzəm, ˈbʊd-] *Buddhism*), [ˈtrɒtskiˌaɪt] *Trotskyite* (≠ [ˈsʌlfəɪt] *sulfite*), [ˈkrɪtɪˌsaɪz] *criticize* (but [ˈbæptəɪz] *baptize*, in addition to [ˈbæptɪˌhaɪz]), [ˈskɒləˌʃɪp] *scholarship* (≠ [ˈfrɛnd(ɔ)ʃɪp] *friendship*), [ˈmɒləˌkjʊːt] *molecule* (≠ [ˈɡlɒbjʊːt] *globule*).

Sentence stress (mostly with English examples)

12.26. It is advisable to consider as *sentence stress* (or *ictus*), every case of word stress which remains stressed in sentence context, and does not become reduced. Generally, English does not reduce its ictuses; as a matter of fact, we can easily have examples such as the following (please, note that in phonotonic transcriptions, the symbols /, ; ./ indicate intonation, not just a separation of example words, together with /?/ and /ɛ̃ ; ɪ/, which are less ambiguous at first): [ðɛnː ˈθɪriː næs ˈblæk ˈkæt̩s ˌræn ˈaʊt.] *Then three nice black cats ran out.*

It is preferable to avoid using the term 'sentence stress' to refer to the sentence *focus*; this last notion refers to the word, or words (and therefore concepts), which in a given utterance are communicatively more *prominent*; in fact, they are highlighted by virtue of being *new* to the conversation (as opposed to being already *given*, or *known*).

Sentence stress and focus are in fact two distinct attributes, although they are not necessarily incompatible. In fact, they can both be present in the last stress group, even though this possibility is statistically the least frequent: [aɪˈnevə ˈseɪd ˈðæt wəz ˈtʃɪruː.] *I never said that was true.*

In practice, it is much more probable for the sentences above to be uttered as follows (although we leave with readers the task of making their own phonotonic transcriptions): [aɪˈnevə ˈseɪd ˈðæt wəz ˈtʃɪruː.], or [aɪˈnevə ˈseɪd ˈðæt wəz ˈtʃɪruː.], or better [aɪˈnevə ˈseɪd ˌðæt wəz ˈtʃɪruː.] *I never said that was true.*

Therefore, a concrete utterance (which is sufficiently long) will have multiple *ictuses* – *ie* *protonic* syllables and one or more *tonic* syllables (in the rigorous sense of *stressed syllables*, respectively, in the *protune* or in the *tune*).

At the same time, the utterance can even have one or more points which are communicatively *highlighted* (*ie* the *sentence foci* [ˈfɔsə, -kə, -ki]), and these are generally expressed by different proportions of stress and pitch.

The sentence *These are the new co-workers of my neighbor Roberta* (British spell-

ing *-our*) can be variously realized, with single or multiple highlights. We can therefore encounter: [ðiizəðə'njɪu·'kʰzɔwɜ'kɪz·əvməθ'nɛɪbə ɹə'bɜ'ʃe..], or also [ðiizəðə'njɪu 'kʰzɔwɜ'kɪz·əvməθ'nɛɪbɛ· ɹə'bɜ'ʃe..], or else also [ðiizəðə'njɪu·'kʰzɔwɜ'kɪz·əvməθ'nɛɪbɛ· ɹə'bɜ'ʃe..] *These are the new co-workers of my neighbor Roberta.*

In any case, the elements highlighted can be grammemes, as well, in case of particular contrasts. With the examples above, we can have: [ðrɪz·əðə'njɪu 'kʰzɔwɜ'kɪz əvməθ'nɛɪbə ɹə'bɜ'ʃe..], or [ðrɪz'ɑ: ðə'njɪu 'kʰzɔwɜ'kɪz əvməθ'nɛɪbə ɹə'bɜ'ʃe..] (with *are* highlighted), or even [ðiizəðə'njɪu'kʰzɔwɜ'kɪzəv'məθ·'nɛɪbə ɹə'bɜ'ʃe..] (with *new* destressed, but with *my* highlighted, for some particular reason).

Some kind of attenuation can occur in parts of the sentence rendered 'parenthetical', as in the following example, which is spoken as a sort of afterthought: [ɪəvməθ'nɛɪbə ɹə'bɜ'ʃe..] *of my neighbor Roberta.*

12.27. In *idiomatic use*, we find that given word sequences, which can also occur in their literal sense, present outwardly 'strange' (or *marked*) stressing. In fact, grammemes, or qualifiers, are brought out instead of the lexemes that accompany them, because these last are destressed (here shown by means of /|/, which becomes distinctive).

More often, we find the sequence *grammeme + attenuated lexeme*, as in: [fɪ'wɛn:θɪŋ] *for one thing*, [ɒnðɪ'vðə'hænd] *on the other hand*, [bə'nɜ:mɪnz] *by no means*, [bə'ɔ:ɪ'mɪnz] *by all means*, [ə'fɛni,ɹɛɪt] *at any rate*, [ɪ'nɛni,kheɪs] *in any case*, [ɪ'nɛni,əvɛnt] *in any event*, [ɪ'nðæɪ,əvɛnt] *in that event*, [ɒn'ðæt,skɔ:] *on that score*, [ɪ'nɜ:ɔ,ɪ'hæm] *in no time*, [ɪnðə'lɒŋ,ɹɛn] *in the long run*, [(ɪvən) ətðə'bestəv-ɪ'hæmz] (*even*) *at the best of times*.

Some other times, instead, we find *attenuated lexeme + accentuated grammeme*, as in: [ɪ'ə ə'nɜ'wɔɪt] *I know what*, [ɪ'ðæt ə'baɔɪ'tɪ] *that's about it*, [fænsɪ'ðæt] *fancy that!*

Thus, *idiomatic stressing* gives a particular meaning, to certain lexical collocations, which is *not* literally predictable. On the contrary, in the *literal sense*, their stress is the 'normal' one, which *is* predictable from the syntactic order of their words, each one bearing its usual meaning, as for: [ɪ'nðæɪ ɪ'vɛnt] *in that event*, [bə'ɔ:ɪ'mɪnz] *by all means*, [ɒnðɪ'vðə 'hænd] *on the other hand*, [ðæt ə'baɔɪ'tɪ] *that's about it*.



13. Macrostructures

(mostly intonation)

Prominence

13.1. The degree in which a syllable stands out among adjoining syllables in an utterance is defined **PROMINENCE**, as already mentioned.

It is the result of different combinations –depending on languages and speakers– of four fundamental elements: the **TIMBRE** of the phones, which form a syllable (ie their relative intrinsic *perceptibility*, determined by articulatory characteristics), **STRESS**, or *strength* of realization, relative **PITCH**, and **LENGTH** (or *duration*). For practical purposes, it is better to analyze the four elements separately (cf \mathfrak{C} 12).

Rhythm & rhythm groups

13.2. Every language has its own particular rhythm, deriving from the structures of its syllables and rhythm groups.

RHYTHM is the result of regular occurrences of prominent syllables in the speech chain. Generally, the alternation of stressed and unstressed syllables is fundamental. In many languages, stressed syllables may also be long(er) \% or be on a marked pitch – high or low (ie different from the mid –unmarked– value).

To all this, some languages may also add a considerable reduction in the duration \% or timbre of the phones of unstressed syllables (indeed some of these phones, often, drop completely, as happens in English, for instance). In quite a large number of languages, duration \% or pitch do not depend on stress; consequently, they may contribute to increase or diminish the prominence of both stressed and unstressed syllables. These characteristics are *chronemic* (ie short phones functionally opposed to long ones, being phonemes) and *tonemic* (ie syllables with different –functional– tones, ie tonemes).

13.3. **RHYTHM GROUPS** (or stress groups) are formed by –at least– one syllable with strong stress. Usually, a stressed syllable is accompanied by other syllables, with *secondary* (or *mid*), *weak*, or *reduced* (or *weakened*) stresses. Furthermore, they show considerable internal cohesion, not only on the phonetic and prosodic level, but also on the semantic one; which means that they have a precise global meaning.

The rhythm groups of certain languages may have only few weak syllables, alternated with mid ones. Other languages, instead, may have quite long sequences of weak or weakened (or reduced) syllables (with shortened phones and attenuated timbres).

English and the Apulian dialect of Bari and Foggia (and others in higher-south-

ern Italy) belong to the latter group of languages; on the contrary, the following belong to the former: Italian, Spanish, French, Brazilian Portuguese, Polish, Czech, Hungarian, Swahili, Hausa, Japanese, Vietnamese. Other languages have intermediate positions: more or less near one group or the other. For instance: German, (Lusitanian) European Portuguese, Russian, Arabic, Hindi, and Chinese are among them.

Here are some examples: *And there was a large crowd of people waiting for them*: [ɪðwɜzəˈlɑ:ðʒ 'kɦɪɑoð ɣ'pɦɪɪpʃ: 'wɛɪɪɪŋfəðm̩..]; *I'm gonna take all of them to the performances*: [æmɡwɒnəˈɦɛɪk 'sɪləðm̩: ɪððəpə'fɔ:mənsɪz..]. English examples will be given only in neutral British pronunciation, to make comparisons easier and simpler.

On the other hand, let us consider these Italian examples (which are a translation of the English ones): *E c'era una gran quantità di gente ad aspettarli* [etʃ,ʃɛrauˈnaɡraŋ kwanti'tad di'ʒɛn:te: aɖaspettarli.]; *Li porterò tutti quanti agli spettacoli* [li:portetɔt tutti'kwanti: aʎ,ʎispettakoli.].

Pauses

13.4. A PAUSE is a momentary break in speech, which takes place for various reasons: physiological (*ie* breathing), semantic (*ie* meaning), logical (*ie* connection), psychological (*ie* attitudes), and pragmatic (*ie* communicative strategies). It is convenient to measure pauses in reference to the number of syllables which could fill the time of their duration, resulting from an average of both stressed and unstressed syllables. Therefore, we speak of *short*, *mid*, and *long* pauses, of about 3, 6, and 9 syllables respectively, or rather of about 2-4, 5-7, 8-10 mean phonosyllables: [I, II, III].

If a short pause is not certain, or may be missing, it is indicated by [:] and is better defined as a *potential* pause. Sometimes, above all for psycholinguistic or behavioral purposes, it may be necessary to indicate pauses in a more precise way: in hundredths of a second; especially when they differ from expected 'normal' ones.

Usually, in *pause groups* (*ie* what is included between two pauses), normal speech uses two or more rhythm groups, which are linked to convey a fuller sense.

Sometimes, a pause group coincides with one rhythm group; at other times, combinations of rhythm groups in different pause groups change the meaning of similar utterances, often much more than the presence or absence of tones within the utterance itself, as in: *You may go now* [jəmeɪ'gɔ:.. ɪ'na:ɔ:..], quite different from *You may: go now* [jə'mɛɪ.. ɪ'gɔ:ɔ 'na:ɔ:..], again different from *You may. Go, now* [jə'mɛɪ.. ɪ'gɔ:ɔ:.'na:ɔ:..].

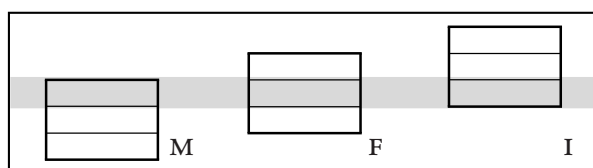
Pitch & tunings (or intonation groups)

13.5. fig 13.1 schematically shows the relationship between the three fundamental types of voices: *male* (M), *female* (F) and *infant* (I); the grey band helps to

realize that the same *absolute* pitch corresponds to quite different *relative* pitch levels. Of course, among the three groups (already introduced in 12.17 & § 12.20) there is a fairly gradual transition, since –for each one– we can easily find more or less high/low voices, in addition to those representing the average of each group.

When two or more pause groups are linked together, their meanings are co-ordinated as well; therefore, they are combined into something wider and more coherent, thanks to a particular intonation contour. Thus we get TUNINGS, which may even coincide with one rhythm group, or one word, possibly formed by one syllable, again: *Yes?* or *Here*.

fig 13.1. Relationship between male, female, and infant voices.



Paragraph & text

13.6. When one or more speakers continue on the same subject, with semantic cohesion, a sequence of tunings is technically called a PARAGRAPH. From a prosodic point of view, a paragraph is usually characterized by given rhythm and intonation features, which determine their internal cohesion, in contrast to other paragraphs, within the same *text*.

Generally, a paragraph ends with a greater pitch lowering (compared to normal pitch), which is marked with [°] at the end. Likewise, a paragraph may begin at a slightly higher pitch, marked by [°] at the beginning. This guarantees internal unity and coherence, in contrast with other paragraphs.

A simple kind of paragraph is constituted by sayings: *When the cat's away, the mice will play* [°wɛn ðw'kxæts ɹwɛɪ. ðə'mæst 'phlɛɪ..°].

A TEXT may be constituted by a speech, a lesson, a (university or public) lecture, a news bulletin, a sermon, a soliloquy, a joke, &c. A *paragraph* may be constituted by the sentences of different speakers, when the text (presenting semantic and pragmatic cohesion) is a conversation, a phone call, an oral examination, an interview, a quarrel, &c.

The text is not necessarily long: even *«Here? – Yes»* may be a text, supposing that the two speakers share certain presuppositions.

Rate

13.7. Languages (and speakers) have different rates of uttering. RATE can be measured in words per minute (100-200 on average), in phono-syllables per second (2-5), or in phones per second (6-20).

In general the number of words and syllables varies according to their structures and extensions; the number of phones varies according to their (phonetic and phonemic) duration.

Moreover, rate varies according to particular semantic, social, and pragmatic factors. *Conversation* itself can be classified in at least three different types: *slow*, *normal*, and *quick*. Consequently, the numbers given above tend to move towards the limits indicated, or even to slightly exceed them.

Pause incidence is connected with rate, too. Indeed, there is a limit beyond which phones cannot be shortened or lengthened without becoming incomprehensible, or ridiculously intolerable. Therefore, the quicker/slower the rate is, the more the duration and number of pauses will be reduced/increased.

In a *normal conversation*, pauses take almost *a quarter* of the total duration of a text. But the time taken by pauses can be longer: up to *half* of the total duration. There are cases (or particular moments) where pauses can even take three quarters of the total time (without falling into pathology); but such cases fall within the aims of paraphonic analysis.

Intonation

13.8. INTONATION (as can already be inferred from § 6.4.5.1-4) is constituted by the relative pitch of syllables forming more or less long sequences of connected speech.

These sequences are called TUNINGS and can consist of pause groups (which, in turn, consist of rhythm groups); but they can also consist in a single word – which can even be monosyllabic: *No*. – *No?* – *No!* – *No...*

What is essential is that pitch –through given differences– adds (or, rather, gives) different pragmasemantic nuances –such as ‘statement, question, command’, &c– to phonic sequences which could otherwise be identical.

Thus the difference obtained is not merely semantic, conceptual, as in the case of ton(em)e languages, such as Chinese. However, by using the same principles and the same symbols of syllabic-tone notation, we can accurately (and without too many problems) transcribe the characteristics of pitch and strength of the syllables of a whole utterance. In fact, stress-tonal signs show both the relative pitch

fig 13.2.1. ‘Icono-tono-graphic’ examples for intonation.

1		<i>See you on Saturday.</i>
2	(<i>Will they</i>)	<i>see you on Saturday?</i>
3	(<i>If they don't</i>)	<i>see you on Saturday... (it'll be a total disaster.)</i>
4	(<i>If they don't</i>)	<i>see you on Saturday... (don't worry about it.)</i>

We must make it clear at once that *written* sentences are one thing, while the *spoken* language is quite another reality, often very different indeed. Naturally, in the spoken language, tunes are much more numerous than ‘simple sentences’ of grammar and syntax, as will be seen below. Indeed, plain and linear syntactic sentences, from a true linguistic point of view (ie actually spoken sentences in real and spontaneous speech, even if uttered by professional actors), very rarely coincide with phonic sentences. As a matter of fact, almost any syntactic sentence is made up of at least two or three phonic sentences, or better *tunings*.

But let us consider *tunes*. Generally they are formed by the *tonic* syllable (ie the stressed one, which is also the last strong syllable in an utterance, in a sense), the *pretonic* (ie the possible unstressed syllable before it), and the *post-tonic* syllables (ie the possible unstressed syllables after it). In the tonetic diagrams (or rather *tonograms*), two post-tonic syllables are indicated (ie internal and terminal); sometimes it is useful to refer to one of them, clearly, in order to highlight typical movements more clearly, above all to distinguish interrogative tunes of the rising type ([·'·]), from those of the falling type ([·'·]). In any case, the term *post-tune* may be used to refer to both syllables, collectively.

We will now consider, concisely (and by looking closely at fig 13.4), the three marked tunes (of neutral British English): *conclusive* (./), *interrogative* (/?/), *suspensive* (/;/), and the unmarked: *continuative* (./) – *On Saturday* [ɒn'sætʃədɛɪ., -ədɪ..], *On Saturday?* [ɔɪn'sætʃədɛɪ., -ədɪ·], (*If not*) *on Saturday...* (*then...*) [ɒn'sætʃədɛɪ., -ədɪ·], (*Perhaps*) *on Saturday*, (*but...*) [ɒn'sætʃədɛɪ., -ədɪ·].

13.10. The best way of dealing with the intonation of a language consists in presenting its structures through appropriate and clear diagrams (ie tonograms), with clear examples and a simple and sufficiently complete notational system (not cumbersome and useless).

First of all, we must repeat that the use and choice of intonation patterns do not depend on syntax at all, but on *semantics* and *pragmatics*, and above all on *communicative goals*. In fact, even if the syntactic formulation is, in the end, the most evident linguistic rendering (for those who are used to reading and writing), in actual fact, it is nothing but a faithful representation of the pragma-semantic way to express concepts and thoughts, peculiar to every language.

If, for instance, we write –and beforehand say– *I've been looking for this for ages* [ɪəvbiŋ'ləʊkɪŋ fəðɪs.. fɪ'leɪdʒɪz..], the superficial formulation at hand is only the inevitable result of the mental and linguistic processes that produce, in English, the sentence just seen, although with slight possible variations.

In actual fact, it results from the juxtaposition of different concepts (each one indicated by /./, or [·'..], or (<')) in a single syntactic string, seemingly simple and straightforward, but actually very complex, as is obvious from its prosodic structure, if supported by an appropriate intonation pattern, as indicated by the small but precious signs used.

Let us now examine the intonation structure of neutral British English. However, we must first consider a general scheme, which will enable us to really *see* its characteristics. Thus fig 13.2 gives the diagram of tunings. It shows the use we make

–when speaking normally– of pitch heights of the various syllables forming the different possible utterances in a given language.

In order to make experiments on the pitch of different syllables, it is certainly better to use identical vowel phones, to avoid being mistaken by articulatory timbres, possibly interpreted as differences in pitch. This is easier in languages like Italian, where examples like *Rimini, concorrono, Anna assaggiava quaranta ananas* [ri:mini, kon'kor:rono, 'anna assadʒ'dʒa'va kwa'ran'ta 'a:nanas] are not difficult to find (despite their meanings, 'the name of an Italian town, they are in competition, Ann tasted 40 pineapples'). As already said, these identical vowels help not to be biased, or misled, by different timbres; since front *V* seem to have a higher pitch, compared to back ones, which give the opposite impression.

Tunings (or intonation groups)

13.11. Tunings consist (as already seen, cf § 13.8-9 & fig 13.2.2) of a protune (in our example *I am transcribing the following example* [æm'tʃɹɪbɪŋ ðə'fɒl-əwɪŋ ɪg'zɑ:mpɫ]) and a tune (*phonetically* [fə'netɪkli.]). In this case, we have a normal protune and a conclusive tune. The latter is represented, tonemically (in an abstract, more theoretical, way) by /./, and tonetically (in a more realistic way) by [·'..] (or by ⟨'⟩, in a graphemic text).

The number of syllables in the example has been calculated on purpose in order to have full correspondence between the tonogram and the syllables of the sentence, to be able to show the characteristics more clearly.

Of course, in normal speech, it is unlikely to find sentences with the same number of syllables; however, the usefulness of the diagram is not compromised, since the actual syllables available (whether more or less than 14) share pitch heights in a fair way. So they may either compress the movement of several syllables into only one or two, or expand it over a larger number of syllables: *Yes, we do* or *Our aim is to pass on ideas, techniques, and practical activities, which we know work in the classroom* (even if this last example, more realistically, will be divided into more parts, with the addition of the respective tunes, mostly continuative), thus: *Our aim is to pass on ideas, techniques, and practical activities, which we know work in the classroom*. In a phono-tonetic transcription, we have: [ɑ:ɹ̄'ɛɪm ɪz̄'pʰɑ:s 'bɪn. əə'dɪvɪz̄. ʰhek'ni:ks̄. əm'pʰɪæktɪkɫ̄ æk'tɪvɪz̄ɪz̄. wɪtʃwɪ'nɜ:ɔ 'wɜ:k̄. ɪnðw'khlɑ:s̄. ɹ̄ɔm..].

Protunes

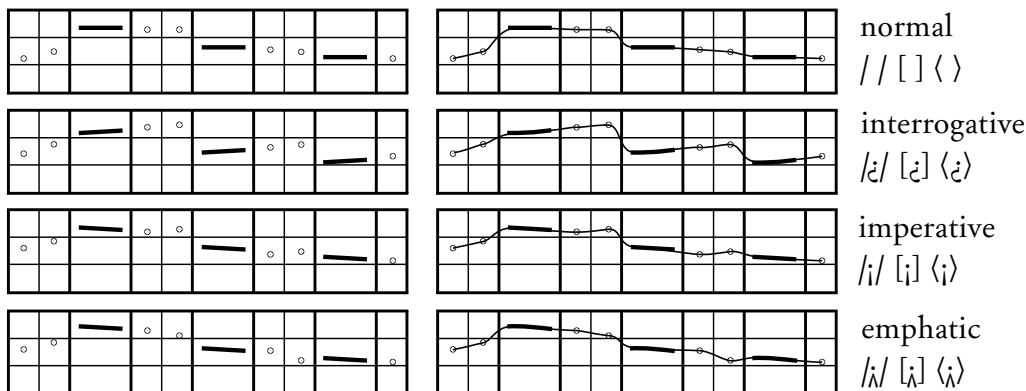
13.12. fig 13.3 shows the four protunes (of neutral British English): one is unmarked, or *normal*, and has no symbol; three are marked: *interrogative* |ɿ/ [ɿ], *imperative* |i/ [i] (for instance: *Pay attention!* [ɪ'pʰɛɪ ə'tʰɛnʃn̄.]), and *emphatic* |λ/ [λ] (*We have to check everything very carefully!* [λwɪ'hæv̄ ʰɛ'tʃhek̄ λ'ɛvɪθɪŋ. λ'vɛɹɪ'kheɪfɪ.]).

fig 13.3 shows, on the left, sketchy tonograms; on the right, they are given in a

more realistic way. Actually, the schematic diagrams are sufficient indeed, since these tonograms necessarily generalize and normalize the data, allowing slight differences of realization, as well.

On the contrary, for teaching and learning purposes, these schematic tonograms are decidedly more useful, making comparisons with those of other languages possible.

fig 13.3. Protunes of neutral British English.



Tunes

13.13. fig 13.4 shows the three marked tunes (of neutral British pronunciation, again both schematically and realistically) – *conclusive* /./ [·' ..] (or (')), should we want or need to use them inserted in current orthography, instead of in transcriptions), *interrogative* /?/ [·' ·:] ((')), and *suspensive* /;/ [·' ·:] ((,)) – in addition to the unmarked one, *continuative* /,/ [·' ·] ((')).

The marked tunes have a functional charge, which is crucial for communication, as they oppose one another distinctively. The unmarked tune – the continuative one – may be considered as the neutralization of the three marked ones (since each of them would be inappropriate in certain – less important – contexts, being too specific and having very definite functions).

The aim of the continuative tune is, above all, to oppose a theoretical 'zero' tune. It is quite different from a straightforward and progressive flow of enunciation, without the slightest variations (or breaks), even theoretical or potential. Its only purpose is to slightly highlight a word, compared to a complete non-occurrence of tunes (as happens within a protune).

Indeed, there is a difference between *I 'saw 'six 'men* [aə̃'sə: 'sɪks 'mɛñ..] and *I 'saw 'six 'men* [aə̃'sə: 'sɪks· 'mɛñ..]; in the latter, of course, *six* is more prominent than in the former, since it has its own tune, instead of being a part of the same protune.

At the end of § 13.11, we have seen that a syntactic string does not generally correspond to just one tune; in fact, more or less numerous continuative tunes occur, otherwise the sentence would not sound spontaneous and convincing. At first, one does not fully realize this internal subdivision, which is completely natural. Its ap-

appropriate use goes entirely unnoticed; whereas, its absence would not pass unnoticed at all (as happens in unprofessional reading or recitation).

For instance, if we consider an utterance such as *Look! the imprints of a bear*, we realize that it can be said in many ways – apart from actual and paraphonic considerations such as the *fright* taken at the sight, or the *delight* expressed by naturalists, or the *satisfaction* felt by hideous poachers... (all of them are rendered with different nuances, clear and easy to interpret). Of course, this is different from a unitary sentence such as *Look at the imprints of a bear*, in just one tuning: [ˈlɒkwət ðiˈɪmpɹɪnts əvəˈbeɪ̃ə..].

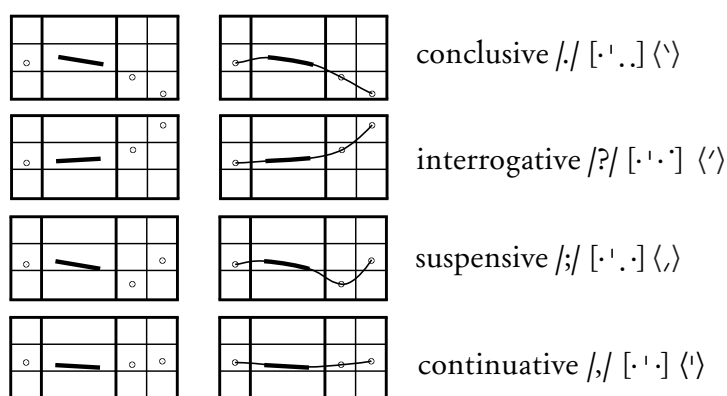
13.14. Thus, if we go back to the original utterance, what we find is something closer to a natural exposition: [ˈlɒk.. ðiˈɪmpɹɪnts əvəˈbeɪ̃ə..]; in fact, in the same sentence, there are two pragmatic concepts: the imprints and its sighting.

If we then divide it into three parts (of course, with three tunes), the nuances expressed are more detailed: [ˈlɒk.. ðiˈɪmpɹɪnts.. əvəˈbeɪ̃ə..]; in this way, we can manage to separate, conceptually too, imprints of different shapes.

After all, it is possible to use some continuative tunes (ie unmarked /,/ as already seen in the previous section), and this will add something to elocution (in opposition to a unitary utterance, although this is not for emphasis, of course). It is only a way to make enunciation a little more effective and natural: [ˈlɒkwət· ðiˈɪmpɹɪnts· əvəˈbeɪ̃ə..] (and variations).

By considering an example like *You must read further books on this particular subject*, again, we can easily see that there are several ways of saying it. Apart from a quite flat realization in a single tuning, as: [jəməs̄ˌɹɪd̄ ˈfɜːðə ˈbʊks ɒnðɪspəˈtɪkʃjələ ˈsɛbdʒɪkʃjələ..], we can have: [jəməs̄ˌɹɪd̄ ˈfɜːðə ˈbʊks· ɒnðɪspəˈtɪkʃjələ ˈsɛbdʒɪkʃjələ..], or: [jəməs̄ˌɹɪd̄· ˈfɜːðə ˈbʊks· ɒnðɪspəˈtɪkʃjələ ˈsɛbdʒɪkʃjələ..], or else: [jəməs̄ˌɹɪd̄· ˈfɜːðə ˈbʊks· ɒnðɪspəˈtɪkʃjələ· ˈsɛbdʒɪkʃjələ..]. We could even have: [jɪm·u· məs̄ˌɹɪd̄· ˈfɜːðə ˈbʊks.. ɒnðɪs· pəˈtɪkʃjələ· ˈsɛbdʒɪkʃjələ..] (with more and more numerous nuances and implications).

fig 13.4. Tunes of neutral British English.



13.15. A *conclusive* tune is necessarily used whenever a given concept is completed in the speaker's mind. Thus, besides the words which form the sentences, it concerns communicative functions as well, as if, in saying *It's raining cats and dogs*,

we added 'I am stating' – so: *It's raining cats and dogs* [ʊts̄ːɹɛɪnɪŋ ˈkæt̩s ənˈdɒɡz̄ː].

Each tune has a specific function: the *interrogative* communicates 'I am asking': *Is it raining cats and dogs?* [ɪz̄ɪt̄ ɹɛɪnɪŋ ˈkæt̩s ənˈdɒɡz̄ː]; the *suspensive* one communicates 'I am underlining': *If it's raining cats and dogs...* (*it's a calamity!*) [ɪf̄ ʊts̄ːɹɛɪnɪŋ ˈkæt̩s ənˈdɒɡz̄ː | (ɪt̄s̄ wɪk wɹ̄læm ət̄i..)]. The *continuative* tune, instead, simply communicates 'I have not finished': *It's raining cats and dogs (but I don't care)* [ʊts̄ːɹɛɪnɪŋ ˈkæt̩s ənˈdɒɡz̄ː (bət̄ ɪ dɒnt̄ ˈkæɹ)].

It is possible to have a series of conclusive tunes: *Yesterday it rained. Today it's raining. Tomorrow it'll pour. I'm sick and tired. I'll go away!* [j̄ɛst̄ədeɪ ʊt̄ɹɛɪnd̄.. | f̄həd̄eɪ ʊts̄ːɹɛɪnɪŋ.. | f̄həˈmɒɹz̄ ʊt̄p̄ɹɔː.. | æm̄sɪk ənˈt̄hæəd̄.. | æt̄ˈgɔː wɹ̄weɪ..]. However, a suspensive tune is very likely for *Tomorrow it'll pour* [f̄həˈmɒɹz̄ ʊt̄p̄ɹɔː..].

Too often, current writing (which is not at all sophisticated) uses only commas: *Yesterday it rained, today it is raining, tomorrow it'll pour, I am sick and tired, I'll go away.* Thus, with the guilty complicity of schools, one is led to a kind of 'child-like' reading, which makes people utter things like: [° ̄j̄ɛst̄ədeɪ ʊt̄ɹɛɪnd̄..° f̄həd̄eɪ ʊts̄ːɹɛɪnɪŋ..° f̄həˈmɒɹz̄ ʊt̄p̄ɹɔː..° | æm̄sɪk wɹ̄nˈt̄hæəd̄..° æt̄ˈgɔː wɹ̄weɪ..°]. The small rings show the additional pitch movement which is typical of 'bookish intonation', which must be kept well apart from normal (ie conversational) intonation, and also from the typical intonation of text exposition (even done mentally), as we will see in § 13.27.

A further example to show that, normally, writing and punctuation are just miserable devices with syntactical functions, and not at all helpful for reading: *I'm terribly busy: I can't come; I'll let you know; don't be cross* [æm̄ˈt̄ɹɛɪəbli ˈbɪzi.. æˈkɹɔːŋ ˈkɹɛmː.. | æt̄ˈlɛt̄j̄u ˈnɜːw.. | ɪˈlɛt̄ju bɪˈkɹɔːs..]. Also the example *I've been looking for this for ages* [æv̄bɪn ˈlʊkɪŋ f̄ðɪs.. f̄ɹɛɪd̄z̄..] shows this characteristic.

Contrary to what grammars keep on repeating, a *comma* does not necessarily indicate a short pause, as a *semicolon* does not indicate a pause which is half-way between the 'short' one of commas and the 'long' one of *full stops* (as it is absurdly 'prescribed'). However, these are the results achieved by schools, ie sadly rigorous and monotonous pauses, which are not able to convey appropriate meaning to sentences (especially when they are read).

And all those who today abuse punctuation, by omitting it almost completely, will they ever pause?

Questions

13.16. A further important point, to always keep in mind, is that an interrogative tune need *not* be used whenever there is a question mark at the end of a sentence! Unfortunately, this is another real mistake taught in schools.

It is important to clearly distinguish the different kinds of questions. Among the more recurrent and normal types, only *total questions* (or 'yes–no questions') request interrogative tunes. These questions are answered by *Yes* or *No* (or, possibly, by *Perhaps*, *I don't know*, *It depends*, &c); but, above all, they do not include

questions just seen could even be formulated as: *I'd like to know how many languages you speak*, and *You must say who told you that*, and *Tell me how you know*.

However, even if in partial questions the conclusive tune has to be used (which is falling just as in statements), there is certainly some pitch difference (apart from an obvious syntactic one) between a question like *When will they buy a new computer?* and a statement like *When they buy a new computer*. This also occurs in languages with identical syntactic patterns, as in Italian *Quando comprano il nuovo computer?* 'When will they buy a new computer?' and *(Lo copiano) quando comprano il nuovo computer* '(They will copy it) when they buy a new computer'.

The difference is in the protune. As a matter of fact, all kinds of questions have something in common, *ie* the interrogative protune, /ɛ/ [ɛ], which, as can be seen in fig 13.3, has a partially different pitch contour from the one used in normal protunes. This difference consists in modifying the usual pitch movement, through the anticipation of the typical interrogative curve (/ʔ/ [·'·]), which in neutral British English pronunciation is rising – from mid to high pitch.

The anticipation in question, however, does not exhibit the actual change from mid to high, but reproduces it on a small scale, by distributing pitch heights among the stressed and unstressed syllables. Thus, it modifies the usual contour of the unmarked protune only partially.

Nevertheless, this is quite sufficient to make the difference perceptible right from the beginning, *ie* on the very first syllable(s). So, in the Italian examples too, the difference is surely there already on *Quan-*, and increases on *-do* (and so on – often together with a different degree of stress). But the symbol /ɛ/ [ɛ] alone is sufficient to indicate the pitch difference that the ear clearly hears: ɛ₁*Quando* 'comprano il nuovo computer?' [ɛ₁'kwando 'kom:prano. ɛ₁ɛ₁il'nwɔ'vo kom'pjuter.], in comparison with *Quando* 'comprano il nuovo computer' [ɛ₁'kwando'kom:prano. il'nwɔ'vo kom'pjuter.]. Going back to the English example, we have: ɛ⁻*When will they buy a new computer?* [ɛ⁻wɛnɪ̯ ðeɪ'ba:ə. ɛ⁻ə'njɪ'u kʊm'phju:ɹe.] and ⁻*When they buy a new computer* [wɛnðeɪ'ba:ə ə'njɪ'u kʊm'phju:ɹe..].

Besides, as all of the partial questions, these too can be said with a continuative tune (which renders them less categorical), or by attenuating the tune (cf § 13.21-3).

The question about the *computer* already shows that the plan of buying a new machine (hopefully a new Mac) was 'known', or 'given', not a 'new' fact. Equally, we have a parenthesis as soon as an example like ɛ⁻*When are you leaving?* [ɛ⁻wɛn: əjə'li:vɪŋ..] becomes a known fact as far as the departure is concerned: ɛ⁻*When are you leaving?* [ɛ⁻wɛn:.. ɛ⁻əjə'li:vɪŋ.]. Usually, this also happens –for pragmasemantic reasons– in sentences like: ɛ⁻*How much does it cost?* [ɛ⁻haʊ'mʌtʃ.. ɛ⁻dʒɪk'kɒst.]; whereas realizations such as ɛ⁻*How much does it cost?* [ɛ⁻haʊ'mʌtʃ dʒɪk'kɒst.]; are to be found only in 'teaching' recordings (and, unfortunately, what we actually hear, in too many recordings, is ɛ⁻*How much does it cost?* [ɛ⁻haʊ'mʌtʃ dʒɪk'kɒst.], indeed!).

13.19. Let us briefly move back to the kind of intonation used at school, which makes people say *₁[ɛ⁻wɛnɪ̯ ðeɪ'ba:ə ə'njɪ'u kʊm'phju:ɹe.] *ɛ⁻*When will they buy a new computer?* whose meaning, strictly speaking, would be closer to 'Would you mind repeating that? I didn't quite catch what you said. Did you ask about when

they're going to buy a new computer?', ie ξ *When will they buy a new computer??*^o [ξ $\bar{w}en\ddot{t}$ $\delta er'ba\theta$ $\theta'nj\mu'u$ $k\ddot{u}m'phj\mu'u\ddot{t}\epsilon$.^o] (where ^o indicates a higher pitch raising).

In actual fact, there is a big difference, since the classic 'bookish question' (as we will see presently, in § 13.20 & 13.25) is: [$\bar{w}en\ddot{t}$ $\delta er'ba\theta$ $\theta'nj\mu'u$ $k\ddot{u}m'phj\mu'u\ddot{t}\epsilon$.^o]. That is, an interrogative tune is added at the end of conclusive tunes ([...]+[^o]), as if it were not real communication, but rather a kind of drill in order to 'identify' a question, and by 'concluding' it –only at the end– with what is thought to be expected (according to the 'rules', ie with an interrogative tune).

However, this operation goes against the rules of real communication completely; and practically without distinguishing between total and partial questions. Indeed, even a question like [ξ $\bar{w}dts$ $j\sigma''n\epsilon'IM$.^o] *What's your name?* is a classic example, instead of [ξ $\bar{w}dts$ $j\sigma''n\epsilon'IM$.] *What's your name?*

13.20. There is also a difference between: ξ *How many times shall I say that?* [ξ $\bar{h}a\theta,meni$ $'\theta ha'\theta m\ddot{z}$. ξ $\bar{l}a\theta's\epsilon'I,\delta\theta\ddot{t}$.] –a normal partial question– and ξ *How many times shall I say that?!* [ξ $\bar{h}a\theta,meni$ $'\theta ha'\theta m\ddot{z}$. ξ $\bar{l}a\theta's\epsilon'I,\delta\theta\ddot{t}$.] –a (partial) rhetorical question, which certainly does not ask for information about the number of times, but instead communicates a meaning like 'Will you obey me at last?' (please, note the emphatic stresses).

In addition, there can also be a *polite* partial question: ξ *How many times shall I say that?* [ξ $\bar{h}a\theta,meni$ $'\theta ha'\theta m\ddot{z}$. ξ $\bar{l}a\theta's\epsilon'I,\delta\theta\ddot{t}$.], which uses the unmarked continuative tune in order to make the question less brusque, as in ξ *What's the time?* [ξ $\bar{w}dts$ $\delta\theta'\theta ha'\theta m$.], ξ *Who is it?* [ξ $\bar{h}\mu'u$ $'\nu z\ddot{t}$.], decidedly more suitable, above all with strangers, than ξ *What's the time?* [ξ $\bar{w}dts$ $\delta\theta'\theta ha'\theta m$.], ξ *Who is it?* [ξ $\bar{h}\mu'u$ $'\nu z\ddot{t}$.].

All this demonstrates that syntactical punctuation and word order are not at all sufficient to determine which kind of intonation is the most appropriate for a given sentence.

Besides, if people ask ξ *Can you hear me?* [ξ $k\ddot{u}nj\theta'h\theta mi$.], their intention is certainly not to check whether their listener's hearing is (still) good, but rather whether it is possible or not to ask them for something, talking normally.

Obviously, there are many nuances which can be detected in the various kinds of questions that –every day– we can produce or hear. These questions may be participating, polite, inquisitive, formal, detached, ironic, sarcastic, and so on. In all these cases, the paraphonic component, with its varied facets, highly modifies canonical intonation patterns, which are so to say 'expected', producing mixtures of protunes and tunes, too.

Tune modifications

13.21. Even a sentence like *Put it on the table* [$\bar{p}h\theta\ddot{t}\ddot{t}$ $\theta n\delta\theta'\theta\epsilon I b\ddot{t}$.] can be said with different intentions. In fact, *Put it on the table* [$\bar{p}h\theta\ddot{t}\ddot{t}$ $\theta n\delta\theta'\theta\epsilon I b\ddot{t}$.] can sound too brusque and impolite, or too familiar and friendly; these nuances are not explained by syntax, but rather by *pragmasemantics*. Therefore, often ATTENUATION may be introduced, which can be shown by placing an empty dot at mid

height [◦]: *Put it on the 'table*◦ [ˈphɒtɪf ɒnðəˈtʰeɪbɪt̩.◦].

We have seen above that, for repetition (or incredulity) questions, the tune is ACCENTUATED. At the end of a conclusive utterance, and especially at the end of a text paragraph, it is frequent and normal to use the accentuation of the conclusive tune too: *And 'this ,ends our 'programs*◦ [ənˈðɪs ˈɛn:dz ɑːˈphɪʒɔŋjæmz̩.◦].

13.22. Often, again for pragmasemantic reasons, utterances are *specified*, when their communicative highlighting –or sentence highlighting– is not on the final part of an utterance (as generally happens), as we have already seen in some previously presented questions.

For instance, *I've 'bought the ,tickets for the 'concert* [aɪv̩ˈbɔ:t ðəˈtʰɪkɪts fəðw̩ˈkɒnsəʃt̩.] we find a ‘normal’ tuning expected according to the tonogram. However, one may have to say: *I've 'bought the ,tickets for the ,concert* [aɪv̩ˈbɔ:t̩. ɪðəˈtʰɪkɪts fəðw̩ˈkɒnsəʃt̩.] – possibly even with some emphasis: *I've 'bought the ,tickets for the ,concert* [aɪv̩ˈbɔ:t̩. ɪðəˈtʰɪkɪts fəðw̩ˈkɒnsəʃt̩.], as an answer to a rather doubtful question about somebody's efficiency or memory.

From a pragmasemantic point of view, the most likely intonation pattern, for a question like *Have you bought the tickets for the concert?*, is *¿Have you 'bought the ,tickets for the ,concert?* [¿hævjʊˈbɔ:t̩. ɪðəˈtʰɪkɪts fəðw̩ˈkɒnsəʃt̩.], contrary to school performances and, unfortunately, to most recordings in various teaching courses too, which instead foist absurdities such as **¿Have you 'bought the ,tickets for the 'concert?* *[¿hævjʊˈbɔ:t̩. ðəˈtʰɪkɪts fəðw̩ˈkɒnsəʃt̩.]. Strictly speaking, a sentence pronounced like that would really mean something like ‘Why did you buy the tickets for the concert, you silly idiot!’).

13.23.1. The best way to attract (much) attention to what somebody is going to say is to use a *suspensive* tune: *If they 'haven't under,stood I 'really don't ,know ,what to 'do about it!* [ɪfðeɪˈhæv̩n̩t̩ ʌndəˈstʊd̩. ɪˈrɪəli dɒn't̩ ,knəʊ ,wɒt̩ t̩ˈdɒ ʌbʊt̩ ɪt̩!], *And 'when at ,last I 'came ,round the ,corner, they were al-ready there* [ənd̩ ˈwɛn: əp̩ˈlɑ:s̩t̩ aɪˈkʰeɪm̩ ˈlɑ:ənd̩ ðw̩ˈkɒns̩n̩. ɪˈdeɪwɜ:ʃt̩ ˈrɛdi ˈðeɪ.], or to clearly separate the parts of an utterance: *¿Are you going by ,bus, or 'walking?* [¿ɑ:ju ˈgɔ:ɪŋ baɪˈbʌs. ɜ:σˈwɔ:kɪŋ.], or to announce in advance the end of a complete list of items: *Apples, 'pears, ,strawberries, 'cherries* [ˈæpl̩z̩. ˈpiəz̩. ˈstɹɔ:bɜ:z̩. ˈtʃɛrɪz̩.].

A suspensive tune can also occur at the end of particular incomplete utterances: *I 'did ,try...* [aɪˈdɪd̩ ˈtɹaɪ.], *You'll 'see one ,day or a ,nother...* [jʊˈsri wɛnˈdɛɪ σˌtʃəˈnɛðe.]. ‘Intermediate’ degrees are possible: *I 'did ,try...*◦ [aɪˈdɪd̩ ˈtɹaɪ.◦], *You'll 'see one ,day or a ,nother...*◦ [jʊˈsri wɛnˈdɛɪ σˌtʃəˈnɛðe.◦] (with attenuated suspensive tunes); *I 'did 'try*◦ [aɪˈdɪd̩ ˈtɹaɪ.◦], *You'll 'see one ,day or a 'nother...*◦ [jʊˈsri wɛnˈdɛɪ σˌtʃəˈnɛðe.◦] (with attenuated conclusive tunes); and *I 'did 'try* [aɪˈdɪd̩ ˈtɹaɪ.], *You'll 'see one ,day or a 'nother* [jʊˈsri wɛnˈdɛɪ σˌtʃəˈnɛðe.] (with no attenuation of conclusive tunes). Obviously, they present different communicative nuances, which are fairly easy to imagine.

In addition, some particular words can receive emphasis. Here we will not enter the field of paraphonics, which adds further nuances, *ie* emotional (*eg* sadness, shyness, threat...), and social as well (*eg* skill, supremacy, arrogance...). Certainly,

these characteristics are real and present, in actual communication, but they are even more complex to analyze, describe, and transcribe. Unfortunately, too many intonation analyses and notations are still unable to adequately separate these two quite different fields of intonation, although, obviously, in speech they are inseparable, but... distinct! This means that it is important both to develop full awareness and to succeed in using a notation system which is fairly appropriate, but obviously not too simple. However, see \mathcal{G} 14.

13.23.2. In some languages, such as English and French, but also Italian and others, according to communicative aims, when there are some *implications*, quite frequently, a *suspensive* tune may be used (with or without attenuation, or possibly accentuation), or a *continuative* one.

This use is more likely to be found in phrases like the following, in place of the tune one might infer from writing and syntax: *Hi!, I'd like a pizza, Go straight ahead, Make yourself at home, Can I have that chair?* or in French: *Bonjour!, J'aimerais bien une pizza, Allez tout droit, Asseyez-vous, Ce n'est pas possible!* or in Italian: *Ciao!, Vorrei una pizza, Vada sempre diritto, S'accomodi, Posso prendere una sedia?*

Quite often, these two tunes are used –instead of conclusive (or attenuated conclusive) ones– when there is no real planning of what is being said, differently from what actors do with a text they already know and have ‘studied’, just to render it in the best possible way (and there *is* a difference – which is quite easily noticed!).

When people talk with no previous planning, in addition to the task of putting together the things to say, another problem arises: trying to avoid being interrupted by their interlocutors, while trying to manage to collect the ideas they want to present. Thus, using different tunes from the conclusive one, obviously, also has the aim to try to achieve this very result, and at the same time communicating that the speakers have not finished their exposition yet.

13.23.3. Another frequent use of *non-conclusive* tunes derives from the insecurity –or weak conviction– of the speakers about what they are saying, or towards their interlocutors felt to be ‘dominant’, by superiority of prestige, role, age...

Often, it is a real behavioral *implication* –of the speaker, not about the message– which conveys ‘non-invasiveness’, in different mixtures of friendliness, deference, hesitation (precisely with /;/, or even simply /,/).

However, sometimes it is actual invasiveness indeed –although not really aggressiveness– due to an excess of liveliness or talkativeness, which prevent the speaker from taking pauses and almost breathing, with the result that normal conclusive tunes are practically missing (or are very much attenuated). On the contrary, more pauses and more conclusive tunes would give the listener relief.

Indeed, this use of non-conclusive –ie continuative or suspensive– tunes includes reported cases of ‘rising tunes’ above all for variants of English (firstly from New Zealand, Australia, and North of England, but now even for the rest of Great Britain, and for Ireland and North America, too), instead of the expected or predictable conclusive tunes.

Although this phenomenon has been detected and described in the areas just

mentioned, it is nevertheless not absent elsewhere, and for other languages. Neither is it something absolutely new, but simply something which can show itself more freely – alas! – without real drawbacks or excessive social stigma.

However, the problem of these reports (even of ‘high rises’, as in [total] questions) arises from the fact that, still too often, intonation is dealt with according to the old method of the British school of phonetics, which is mainly based on two opposing types: falling and rising (although the British method of intonation had been really innovative and praiseworthy, in the first half of the twentieth century [even compared to other schools and methods, even currently followed by some, as the so-called *ToBI*, which tries to do tonetics by using computers, *ie* using one’s eyes instead of one’s ears).

Unfortunately, the ‘rising’ type included both interrogative tunes (usually: [\cdot^{\cdot}], but also [\cdot^{\cdot}], &c) and suspensive tunes (generally: [\cdot^{\cdot}], or [\cdot^{\cdot}] [\cdot^{\cdot}], or [\cdot^{\cdot}] [\cdot^{\cdot}], and [\cdot^{\cdot}]), which really have rising movements on the post-tonic syllables; but on (quite) different levels. In fact, for /?/ the terminal post-tonic syllable is high indeed, whereas for /;/ it remains within the mid band, as can be clearly seen in our tonograms, including the ‘Oceanian’ ones of Australia and New Zealand. Therefore, they must absolutely not be combined in descriptions (although this is just what is still done, too often, indeed).

The solution to this problem is to adequately separate the ‘linguistic’ level of intonation (*ie* the linguistic system) from the ‘paralinguistic’ one (*ie* *paraphonics*). Of course, machines cannot do that, since even environment noises ‘are a part’ of a sound message for a machine. Rightly, it is up to ‘natural’ phoneticians to define in advance an inventory of tunes and protunes, with their actual realizations, in order to be able to separate them from paraphonic superstructures, which are additional. No doubt, the latter belong to *language usage*, but *not* to intonation proper (or linguistic intonation). Instead they belong to paralinguistic intonation. It is quite obvious that rising post-tonic syllables, if modified by an equally rising superstructure, produces the global, ‘raw’, result consisting in an even more rising movement. However, the analyst’s experience and skill could avoid gross errors such as those of interpreting the pitch movements as if they really belonged to the intonation system of a given language. On the contrary, they are the (natural and inevitable) result of the combined actions of (true) intonation and paraphonics (according to pragmatic, geographic, and sociolinguistic characteristics).

A similar case of communication *or* description misunderstanding occurs when some northern Italians (in particular from the northeastern region of Veneto) use suspensive tunes with post-tonic syllables of the [\cdot^{\cdot}] kind, or similar continuative tunes (with a limited rising movement, *ie* within the mid band, but still of a rising kind). Indeed, people from other regions often interpret such *post*-tunes as if they were [\cdot^{\cdot}], *ie* interrogative, whereas they are nothing of the kind (neither physically, nor intentionally).

fig 13.5 shows the *attenuation* mechanism of marked tunes (we briefly mentioned above), whereas fig 13.6 shows the *accentuation* mechanism of the two tunes which are functionally the most opposing (conclusive and interrogative – already mentioned too).

fig 13.5. Attenuation of tunes.

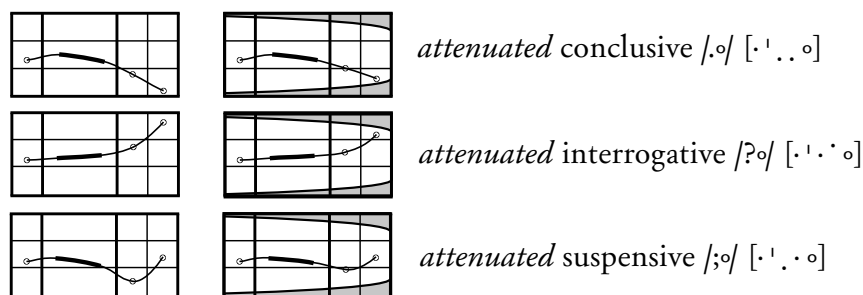
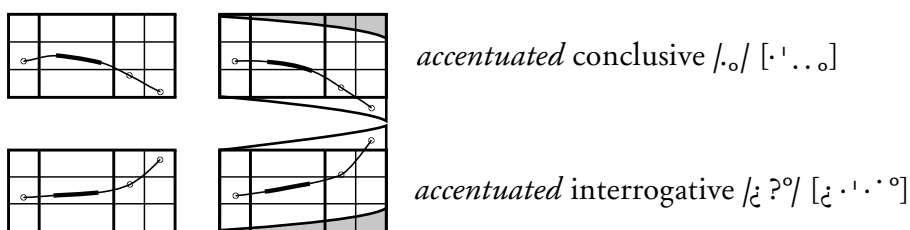


fig 13.6. Accentuation of tunes.



Parentheses & quotations

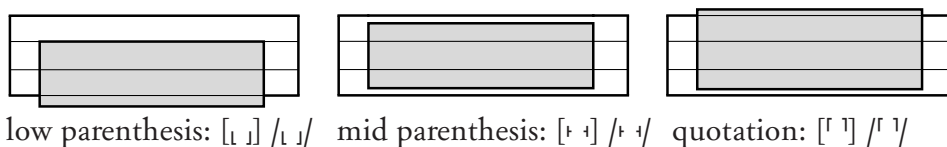
13.24. Lastly, again in a concise way, let us consider fig 13.7 which shows the diagram of *parenthetic phrases* (which can be *low* or *mid*, as we will see presently) and of *quotations*. Here we will illustrate parentheses and quotations: *First of all –he said– let's consider 'natural' phonetics, as it's properly called* [ˈfɜːst əv'ɔːl ˌhiːs'eɪd ˌlets kʌnˈsɪdə ˈnætʃʃl̩.. ˌfə'nɛtɪks.. ɪzɪts ˌpɦɪpəpəli ˈkɦɔːfɪd̩..]

Parentheses are characterized by a reduction of stress, an increase in the rate of speech, and low or mid compression; whereas quotations have an increase in stress, a reduction of rate and a raising of pitch (without compression). Therefore, quotations are –practically– the opposite of parentheses.

In transcriptions, it is neither necessary nor recommendable to try to show these prosodic peculiarities, not even with regard to stress, which remains marked as usual (without notational reductions, [ˌ], or accentuations, [ˈ]). The symbols [ˌ], [ˈ] are more than sufficient to bear in mind all these differences, with respect to 'normal' utterances.

Obviously, quotations must not be confused with 'direct speech' – since, in *First of all –he said– let's consider 'natural' phonetics, as it's properly called*, only *he said* would be excluded, because all the rest (and what may follow) *is* direct speech, indeed.

fig 13.7. Diagrams of parentheses and quotations.



conversation –nor is it a soliloquy– because a superstructure is added which gives the characteristic of a *text exposition*. It is recognized by the fact that it presents mechanical and ‘regular’ pauses (which are never too long), and quite attenuated tunes, but most of all they are ‘completed’ by slight sudden rises from low to mid pitch, which are indicated by adding [_o°] after a tune.

We will now give an example and compare it with real bookish intonation – *Thus, in such cases, one must keep calm, follow the instructions, and think long and hard before acting*: (CONVERSATION) [ðəs· ɪnˈsɛtʃ ˈkheɪsɪz· ˈwɛm məsˈkhiɪp ˈkha:m· ˈfɒlɒ ɔ̃ɪnˈstɪʃkʃnɪz· ɛnˈθɪŋk ˈlɒŋ ʊnˈhɑ:d· bɪfɔ:ʃækʃɪŋ..]; (EXPOSITION) [ðəs· ɪnˈsɛtʃ ˈkheɪsɪz:° ˈwɛm məsˈkhiɪp ˈkha:m:° ˈfɒlɒ ɔ̃ɪnˈstɪʃkʃnɪz:° ɛnˈθɪŋk ˈlɒŋ ʊnˈhɑ:d:° bɪfɔ:ʃækʃɪŋ:°]; (BOOKISH READING) [°ðəs· ɪnˈsɛtʃ ˈkheɪsɪz:° ˈwɛm məsˈkhiɪp ˈkha:m:° ˈfɒlɒ ɔ̃ɪnˈstɪʃkʃnɪz:° ɛnˈθɪŋk ˈlɒŋ ʊnˈhɑ:d:° bɪfɔ:ʃækʃɪŋ:°]. Notice, for *exposition*, the difference in the relative height reached by [_o°] (ie lower than [_o°]), and the quite normal pitch at the beginning (in comparison with [_o°] of bookish intonation).

Furthermore, in bookish intonation, *hesitations* can be frequent, generally realized as longer than usual *unfilled pauses* (cf § 13.4), but often they may become *filled pauses*, with autonomous ‘syllables’: [v, ɛ:, ɜ:, ɜ:, m, hm], or with (paraphonic) drawls (not regional ones such as those typical of the American ‘Deep South’): *but the-en you’ll ha-ave to wai-it...* [bətðEEN juuʰhææy ʃuˈwEEɪɪ]. A well-done exposition will limit hesitations to the most, unless they are ‘intended’ to reach two aims: to draw somebody’s attention to certain points, or to fake spontaneous speech, as when improvising, in order to be more welcome and seem smarter.

13.28. *TV news* presents a [_o° ° ° ° ° ° °] superstructure, which permits one to identify the beginning and the end of every piece of news. Good newsreaders restrict themselves to this superstructure, which is necessary and crucial, while accurately avoiding introducing the excessive internal (sudden) rises, which are typical of bookish intonation. But, too often, the news is unduly and arbitrarily segmented so that its utterances are even distorted, to the point that they can communicate different –or even opposite– senses, as regards the actual meanings of the information intentions.

Furthermore, these pauses are quite mechanical and short (but, above all, different from those of ordinary conversation), whereas, near the end of a piece of news, there is frequently an interruption between the last stress (ie the one on the tonic syllable) and the preceding one. All this happens independently from internal cohesion (which would be necessary among the elements), so much so that even a first name is separated from its family name, a verb from its direct object or adverb, and an adjective from its noun, &c: *...the famous opera by Giuseppe| Verdi°; *the efforts to accelerate the structural reforms| of the economy°; *...with new| proposals°; ...

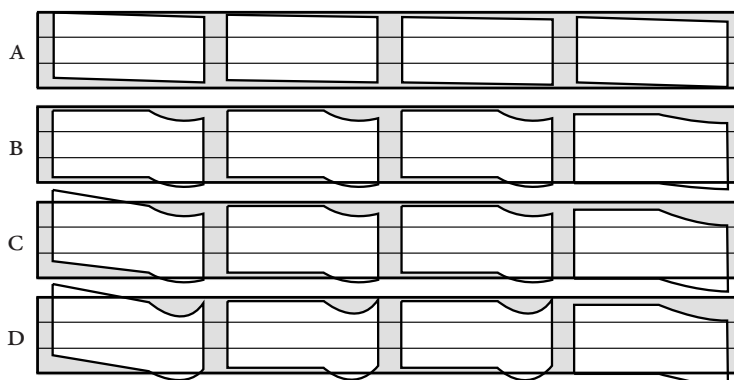
The last examples show that often the normal structure is irrationally distorted; whereas, in previous points of a piece of news, more often distortions regard cases such as: *...the Cannes| Film Festival° – instead of ...the Cannes Film Festival°.

Often, within a piece of news, newsreaders do not keep the end of a sentence

separated from the beginning of the next one: *...they decided to 'meet| in 'London they 'also 'promised... – instead of ...they decided to 'meet in 'London| they 'also 'promised... (ie ...they decided to 'meet in 'London. They 'also 'promised...).

13.29. Obviously, there are many other COMMUNICATIVE ROLES, which must allow people to realize that spoken words are not to be interpreted in a *personal* way, as among friends or acquaintances. On the contrary, it must be clear that they are to be interpreted as a part of a role, thus in an absolutely *impersonal* way, as operator and client. For instance, a ticket collector (*fares, please*), a postal worker (*good morning*), a switchboard operator (*hello, we-are-the-best-in-the-world*), a shop assistant (*good afternoon, Sir, can I help you?*), a stewardess (*flying with us is a pleasure and a guarantee*), a Far-West sheriff (*howdy, stranger!*).

fig 13.8. Paraphonic structures of: conversation (A), exposition (B), TV news (C), and bookish intonation (D).



Their messages (beyond the expected meaning, which is practically superfluous) want to communicate, above all, 'we are just doing our job, and we want you to know'. Therefore, the (ticket) collector, the (postal) worker, and the (switchboard) operator will use paraphonic pitch *compression*: ⟨° / ,/⟩ ⟨° [·]⟩, avoiding using a conclusive tune. Whereas, the (shop) assistant, the stewardess, and the sheriff will make use of paraphonic *raising*, emphatic protunes, and again continuative tunes: ⟨° /_λ ,/⟩ ⟨° [λ ·]⟩ (cf fig 14.1).

After all, even every-day *greetings* among people known by sight, who are not on intimate terms, are just kindnesses, which simply show peaceful coexistence (in comparison to cutting somebody dead); thus greetings can not be either omitted, or too friendly: *good 'morning* ⟨° [gɔb'mɔɹnɪŋ·]⟩, *good 'evening* ⟨° [gɔd'ivnɪŋ·]⟩; in fact, generally, they are reduced to: *'morning* ⟨° ['mɔɹnɪŋ·]⟩, *'evening*: ⟨° ['ivnɪŋ·]⟩, or even to ⟨° [mɔɹŋ·]⟩, and ⟨° [ivŋ·]⟩.

13.30. Let us now also consider '*child-like*' reading. Everybody can easily recall it and recognize it at once (rather annoyingly). However, this is a 'role' again with its conditions. In fact, child-like reading has the typical 'textual' superstructure [° ° ° ° °], with the addition of emphatic *stress* (but without using an emphatic protune) on each expected rhythm group (but also with more subdivisions that often produce two tunes instead of only one).

Obviously, even bookish questions (or *child-like questions*) are included in this pattern. Again they are parts of an utterance which are artificially separated and maintained together only by *textual* superstructures (not by real conversational strategies).

Let us give an example, by comparing the structures of conversations with the superstructures of expositions and *child-like readings* – *To prepare a tasty vegetable soup, the ingredients must be chosen very carefully. What do we need?:* (CONVERSATION) [tʰəpɪl'pʰe:ɜ̃. ə̃tʰeɪstɪ 'vedʒɪbɪʃmʊp. ðiŋ'gɪɪdɪənts. məsbɪtʃhɜ̃ʊzŋ ɪveɪ'kheɜ̃fəli. | ɜ̃wɒt ɔ̃ɪwɪ'nɪɪd̩.]; (EXPOSITION) [tʰəpɪl'pʰe:ɜ̃. ə̃tʰeɪstɪ 'vedʒɪbɪʃmʊp.: ə̃ ðiŋ'gɪɪdɪənts.: ə̃ məsbɪtʃhɜ̃ʊzŋ ɪveɪ'kheɜ̃fəli.: ə̃ | ɜ̃wɒt ɔ̃ɪwɪ'nɪɪd̩.: ə̃]; (CHILD-LIKE READING) [°tʰəpɪl'pʰe:ɜ̃. ə̃tʰeɪstɪ 'vedʒɪbɪʃmʊp.: ə̃ ðiŋ'gɪɪdɪənts.: ə̃ məsbɪtʃhɜ̃ʊzŋ.: ə̃ ɪveɪ'kheɜ̃fəli.: ə̃ | °ɜ̃wɒt ɔ̃ɪwɪ'nɪɪd̩.: °ə̃].

The (alleged) 'remedy' prescribed by schools, in order to avoid the dullness of child-like reading, leads to *flattening* (with pitch compression of the internal parts), to increasing *rate* (with the reduction of many stresses), and to *hypo-segmenting* utterances (with the suppression of many tunes), with a loss of the wished expressiveness: *[°tʰəpɪl'pʰe:ɜ̃ ə̃tʰeɪstɪ 'vedʒɪbɪʃmʊp.: ə̃ ðiŋ'gɪɪdɪənts məsbɪtʃhɜ̃ʊzŋ ɪveɪ'ækjəɪtʃli.: ə̃ | °ɜ̃wɒt ɔ̃ɪwɪ'nɪɪd̩.: °ə̃].

Considerations on intonation

13.31. The fundamental criterion for 'choosing' the appropriate tunes, for each sentence, consists in the *communicative intention* of every single sentence, or sometimes of a part of a sentence, which thus receives a certain tune, often without an actual pause. But the absence of pauses must not make people think that the stream of syllables and words is constant and homogeneous: intonation differences are there (indeed!) in spite of the connection among syllables. Rightly these pitch differences, which are included in the typology of linguistic intonation, convey the nuances of meaning that native speakers instinctively recognize, and thus (re)act accordingly.

Therefore, it is not sufficient to have a *continuative* tune, which is used mainly for subdividing the speech chain into pragmasemantic sequences. These are of fundamental importance to communicate what people think, in order to interact with others. Thus one of the three marked tunes is chosen.

A *conclusive* tune, as we have already seen, adds to the concept being expressed the communicative function of completeness. An *interrogative* one adds the function of request, generally with total questions and, occasionally, with clarifying questions, when people have not (fully) understood or think they can not believe an interlocutor or their own ears.

Lastly, a *suspensive* tune is used to draw attention to what somebody is going to say (or not to say), or to some more or less relevant alternatives.

When people are not slave to syntactic punctuation (if any – or at least syntactically appropriate), in order to obtain satisfactory results, it is sufficient to apply the right *communicative aim* to one's thoughts. However, it is quite obvious that, if people do not *really* know the intonation patterns of their own language (ie in

a perceptive and productive way as well – *not only theoretically*), the result will be a performance which is either regional (for native speakers with no falterings or hesitations), or decidedly foreign (if they try to use the patterns of their own language while speaking another one).

13.32. Since some tunes (and protunes) of certain languages may be very different from those of others, or even opposite (or can remind some opposite functions), it is important to consider the provided tonograms very carefully, to compare them with those of one's own mother tongue (or a regional variant of it), if they are available.

Otherwise, it is highly advisable to endeavor to get them, and try to grasp the differences of the tonograms of other languages or those of the neutral (or 'standard') accent of one's own tongue. If one is able to sing in tune, the operation proves to be better, but this is not at all indispensable: what is essential is to have the will to make this comparison (if one is convinced).

Certainly, a significant number of speakers of tone languages are 'out of tune', and yet they all use their ton(em)es adequately, as for instance Chinese men and women do (and, of course, children as well, if not too young).

Obviously, even less clearcut behavior is to be expected. For instance, a question like ξ 'Who wrote "Romeo and Juliet"?' –instead of the answer 'Shakespeare' [ʃEIK,spɪə..]– might receive a not too sure one, like ξ 'Shakespeare' [ɛʃEIK,spɪə..], or even a far less positive one, like ξ 'Shakespeare...' [ɛʃEIK,spɪə..], where the interrogative protune seeks confirmation.

More convinced answers like λ 'Shakespeare!' [λʃEIK,spɪə..] (ie emphatic) or \imath 'Shakespeare!' [ɪʃEIK,spɪə..] (ie imperative) would still be different from 'Shakespeare' [ʃEIK,spɪə..], in the opposite direction.

These principles hold for all languages, but –obviously– they have to be realized with the actual tonetic characteristics (typical of every single language).

The treatment of the intonation of the various languages (in *HP_r*) might seem to be too sketchy (especially to a hurried reviewer); indeed, it contains all the indispensable elements for the 12 languages dealt with (and their numerous variants too). Instead, in the second part of *NPT*, the *Phonosyntheses* contain only the unmarked protune and the three marked tunes; however, even from these more basic elements, it is quite easy to obtain both the marked protunes and the unmarked tune. This is true both as a general expectation and as an actual work of elicitation.

The three marked protunes are influenced by the three marked tunes, especially on their intertonic syllables. While, the unmarked continuative tune is influenced both by the conclusive and suspensive tunes, being a kind of average between them.

What is important, for intonation, is to bear well in mind all that has been said in this chapter, of course, after having internalized all the motivations and mechanisms.

Structures & generalizations

13.33. The experience obtained with the intonation systems of hundreds of languages allows us to state that the normal, or *basic*, intonation structures are the unmarked protune (/ /) and the three marked tunes (/./ with /?/ and /;/). In addition, we also speak of *fundamental* protunes and tunes: they include the interrogative (/¿/), imperative (/¡/), and emphatic (/λ/) *protunes*.

The three marked protunes announce in advance, in a compressed form (although not necessarily within the mid band, depending on the language), the typical pitch movement of the interrogative (/?/), conclusive (/./) and suspensive (/;/) tunes –respectively– differentiating them from the unmarked, normal one (/ /).

On the other hand, the unmarked, continuative tune (/./) neutralizes the functions –and the tonetic substances, as well– of the three marked ones, producing a compressed movement within the mid band. This movement represents their fusion, by flattening the differences, even though given proportions typical of the original marked tunes are maintained, which contribute towards the differentiation of languages. This remains within the mid band (as can be seen in fig 6.19, fig 6.21, fig 13.4 and fig 13.10 [and in those of the 12 languages of *HPr*]).

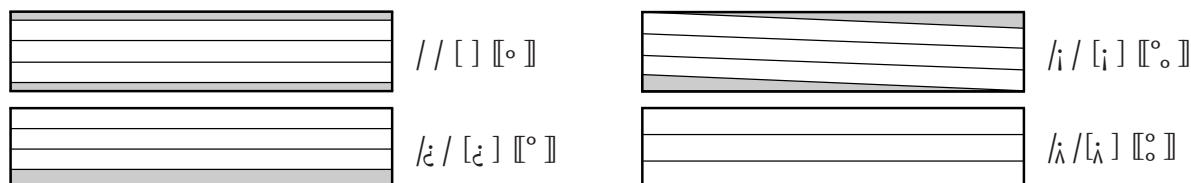
Therefore, the necessary and sufficient notation is simply [[·]·], with a single post-tonic dot, in order to distinguish it –fairly easily– even from suspensive tunes of a mid type, [[·]··], which is normally used in certain languages.

Only occasionally, it may be necessary to exceed an inventory of four protunes and four tunes, as happens in neutral French and certain similar accents (cf § 13.9 and, in *HPr*, \mathfrak{G} 4). For some other languages, the interrogative protune –although remaining only one– is slightly different from the normal type, as happens in Galician, Greek, and Rumanian (cf their phonosyntheses, § 17.1, § 17.53 & § 17.62).

13.34. In *tone languages* (though, more strictly, they should be called *toneme languages*) as Chinese and also Japanese (cf \mathfrak{G} 11-12 of *HPr*, and the phonosyntheses of other tone languages, in \mathfrak{G} 17-22 of *NPT*), *protunes* and *tunes* are subject to modifications with regard to the extent of their pitch range.

As a matter of fact, in general (cf fig 13.9), a normal protune is slightly compressed towards the middle part, / / [[] [°]], whereas an interrogative protune is compressed upwards, /¿ / [[] [°]]. An imperative protune undergoes a certain compression, which starts high, then gradually goes down to end in a low position, /¡ / [[] [°]]. Finally, an emphatic protune does not usually present any compression, remaining completely expanded, /λ / [[] [°]].

fig 13.9. Protunes for tone languages.

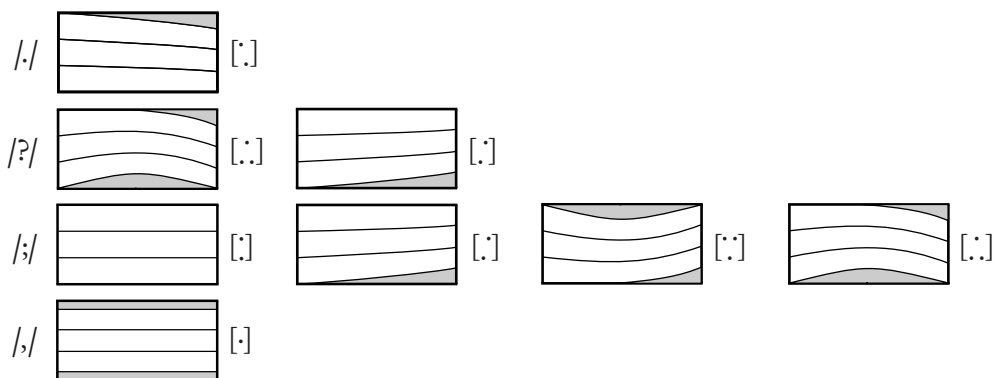


With regard to the tunes of the various tone languages (cf fig 13.10), the conclusive one compresses pitch downwards, gradually, from its beginning (or, as in Japanese, in a more sudden way, at its end).

More often, an interrogative tune is rising-falling, /?/ [.:]; or else rising, /?/ [.:]. The suspensive tune is the most varied one: more often it is expanded (with no particular pitch movements and without compressions), /;/ [.:]; or else it is rising, /;/ [.:]; sometimes it is falling-rising, /;/ [.:]; or more rarely it is rising-falling, /;/ [.:].

As happens with the other languages as well, certain different tonemic functions can have similar tonetic realization, and vice versa.

fig 13.10. Tunes for tone languages.



14. Superstructures

(with ‘paraphonic intonation’, as well)

14.0. ‘Role’ superstructures, or task superstructures, have been introduced, with some examples and classifications, in the sections about intonation, since they are strictly connected with intonation structures (§ 13.25-30). More properly, we might define them as ‘co-structures’; but –to avoid complicating things too much– although maintaining the definition ‘superstructures’ (which are the subject of this chapter), we leave them in the relevant chapter about structures. These co-structures will be presented in a chapter of their own, when they can be studied in greater depth and classified thoroughly. Equally, the chapter on super-structures will (and shall have to) be expanded considerably and more systematically.

Paraphonics

14.1. When the vocal folds (cf fig 14.2) are vibrating along their whole surface, the result is the so-called CHEST VOICE (with actual vibration of the thorax as well); whereas, we get HEAD VOICE if the folds are vibrating only at their sides (thus the thorax has reduced vibration). There are two further quite important positions of the glottis: those which produce *falsetto* and *creaky voice*.

Falsetto, or false voice, is an artificially higher than normal tone of voice, which can be used for expressive, phonostylistic, paraphonic purposes; it is achieved by lengthening the vocal folds more than usual and making them thinner, while they are kept ajar and only the arytenoid cartilages are firmly closed together; therefore they vibrate in a different manner than usual. The larynx itself is deformed a little, and lengthened forwards, as can be seen in fig 14.3. There is little air expenditure, and a general impression of higher pitch and effort than normal. In Hindi, for instance, falsetto is connected with the higher parts of its tunes.

For *creaky voice*, or *laryngealization*, instead, the vocal folds are less tense and less thin than usual, whereas the arytenoids remain firmly closed, so that only the part of the folds not in contact with them are free to vibrate. The folds are completely lax, and produce a series of quick glottal taps and a pitch impression which is lower than usual (fig 14.3).

We can find creaky voice in some languages, in conjunction with low pitch, as at the end of certain conclusive tunes, which are falling, or, as in Mandarin Chinese, in conjunction with the low parts of tunes. In British English, creaky voice on a conclusive tune is considered as upper-class: *Yes, indeed* [ˈjɛs ɪnˈdriːd̥]. Or, in American English, creaky voice is often connected with female speech.

Two other particular phonation types, with paraphonic importance, are *tense voice* and *harsh voice* (fig 14.3).

14.2. By the term PARAPHONICS we mean the particular *or* additional use of phonic (ie articulatory and phonatory) and prosodic elements in current speech. This use is spontaneous (but it can also be fictitious, as good actors demonstrate), and it helps to indicate the speakers' *attitudes* (towards topics or listeners), *emotions* or states of mind (which can be *transitory*, in a particular moment, or *constant/steady*, habitual), and persons' *social role* (in certain different communicative situations).

Generally, paraphonic elements seem to be universal, ie used in every language. Indeed, even without understanding a given foreign language, we can usually intuit, or deduce, the speakers' attitudes, feelings, and social roles. Instead, the *social rules* –more or less unconscious and governing the use of paraphonic elements– are generally very different.

In some cultures, for instance, certain 'negative' feelings must not be expressed, whereas others may be, and 'positive' ones must even be heightened. In other cultures, though, the opposite may be true. Moreover, there are upper and lower limits, changing according to the speakers' age *or* sex, their social roles, and other more or less complex variables.

Here, we shall just explain the characteristics which –alone or combined– contribute to give paraphonic information, when they are superimposed on, or inserted in, the speech chain.

In transcriptions, paraphonic elements are usefully indicated in angled brackets, $\langle \rangle$, in order to identify them clearly and to distinguish them better from prosodic elements, which are more typically linguistic. In fact, the phonic and paraphonic levels should never be mingled, even if, in actual language, they are inextricable.

If one does not do so, the *description of intonation* characteristics becomes not only too heavy, but above all impracticable, and decidedly too subjective. As a matter of fact, even using the same recorded corpus, different phoneticians would inevitably produce different 'data' and descriptions (even if they used computers and acoustic phonetics programs).

Pitch

14.3. Of course, the first paraphonic characteristic to consider is the one that uses pitch in a different way with respect to how intonation uses it. Paraphonically differences can be greater and –what is more– of a fairly different quality, since feelings, moods, and states of mind make use of different pitch extensions.

Usually, in fact, they do not regard every single syllable of an utterance, as in tunings, where pitch is determined by linguistically codified structures (seen in \mathbb{G} 13). Although with possible variations, caused by changes in the factors determining their origin (ie moods, feelings, states of mind, attitudes, and physiological, psychological, and pathological conditions), paraphonic stretches mostly regard whole utterances of different lengths, even from minutes to hours.

With respect to the habitual pitch extent of a given speaker (within the three

categories of voices, ie male, female, and infant), it will be sufficient to get accustomed to recognizing, first of all, the *raising* of general pitch (indicated by ⟨°⟩, between *angled* paraphonic brackets), visually and mnemonically helped by fig 14.1. There the two white lines show the relation with the normal (pitch) extent, which is then remarkably modified depending on the paraphonic use of pitch. We can see that the pitch extent is really ‘raised’ indeed, in comparison with the one considered ‘normal’.

Now, it is sufficient to think of a label such as ‘astonishment’ or ‘aggressiveness’ and say something fitting, like: *And what’s this?* with astonished wonder, or *Don’t be a pain in the neck!* with arrogant aggressiveness. Often, we simply use raising, ⟨°⟩, when we are talking to children, too: *Hello, pretty child, what’s your name?*

With other labels, we can identify the opposite movement to raising, ie *lowering* ⟨◊⟩, as can happen for sadness: *Oh, that’s such a pity!*

We can have simultaneously raising and lowering, too, which leads to *expansion* both upwards and downwards, ⟨°◊⟩, by broadening the three pitch bands, as in cheerfulness: *Here come our dear friends!* Practically the opposite of expansion is *compression*, ⟨◊◊⟩, which is typical of labels like sleepiness: *I could sleep like a log!*

fig 14.1. Paraphonic characteristics of pitch.

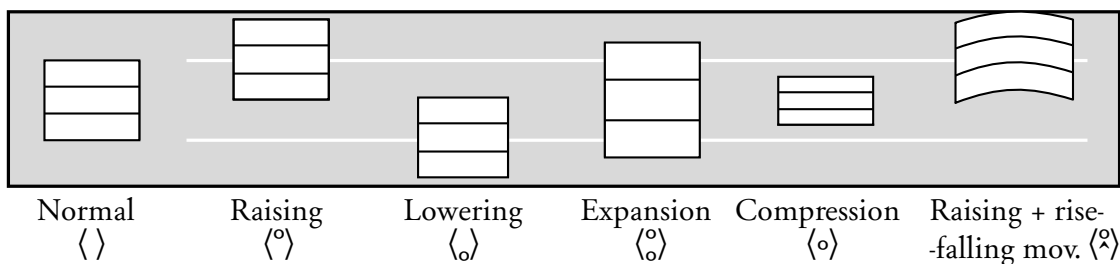


fig 14.2. Normal and paraphonic characteristics of phonation types.

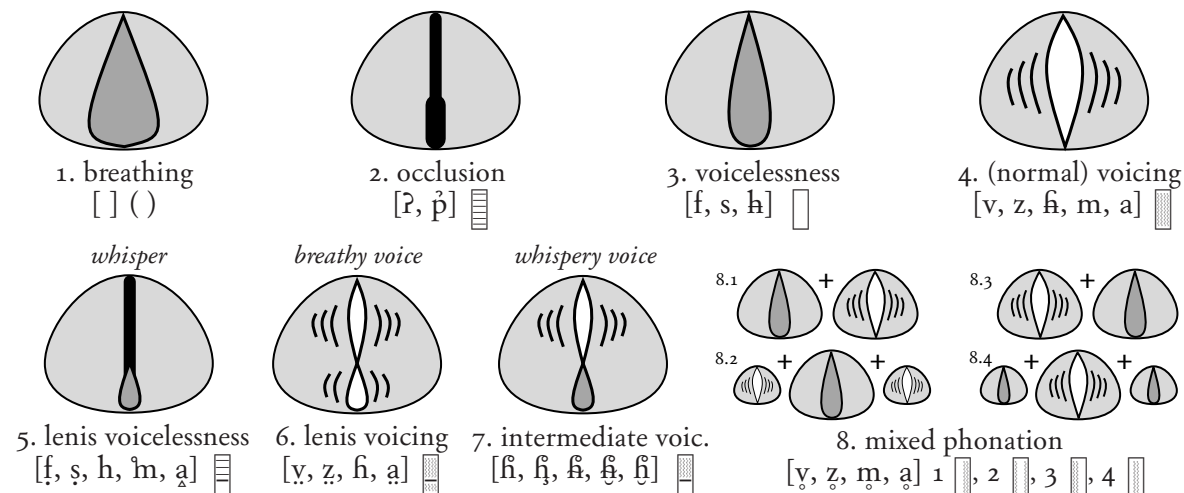
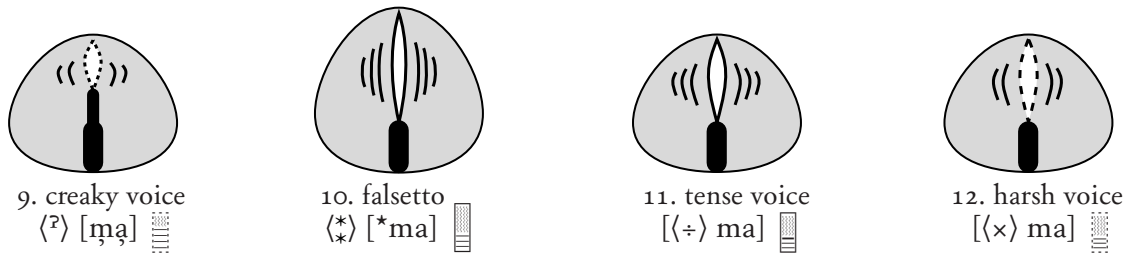


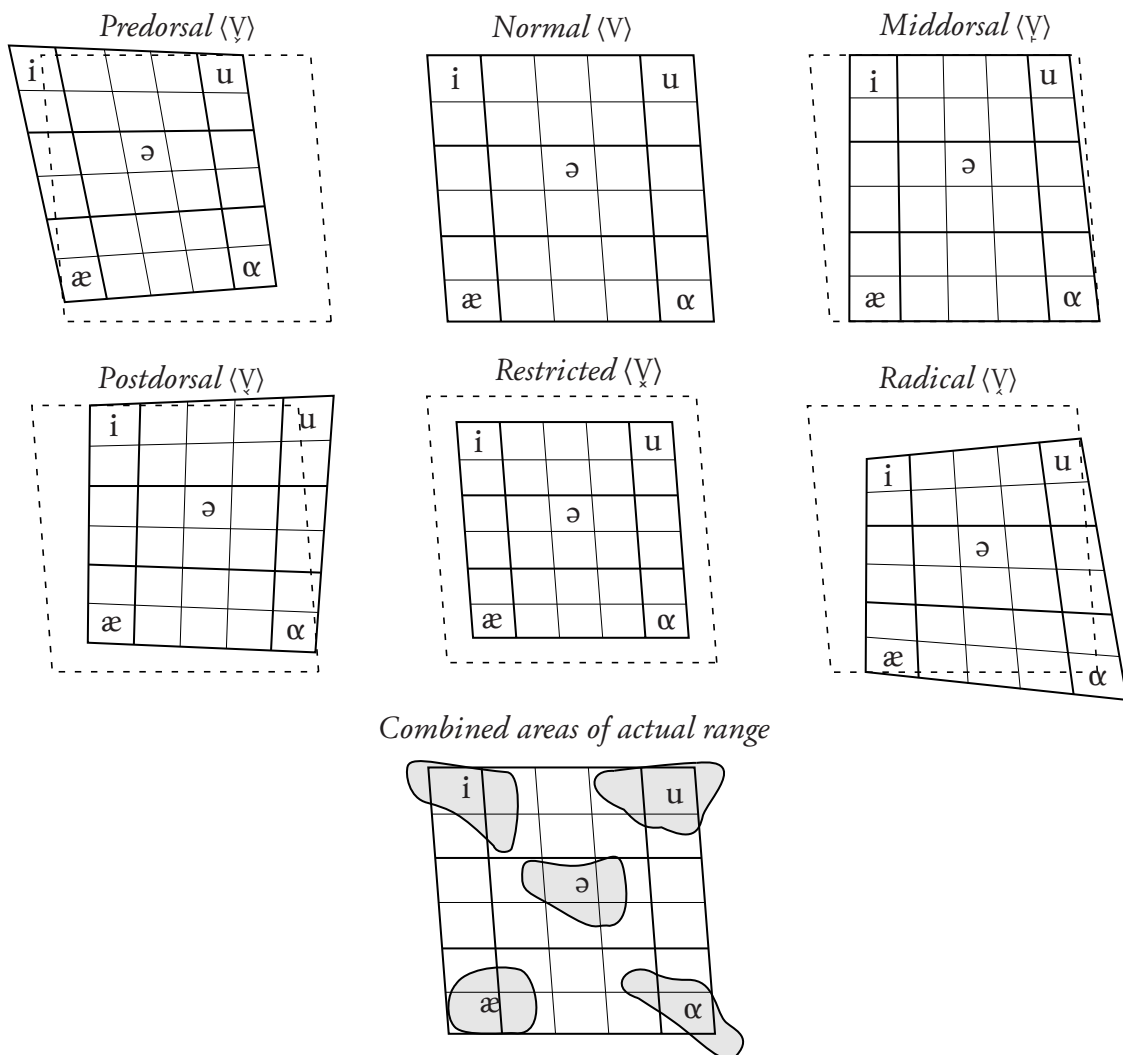
fig 14.3. Some voice qualities, frequent in paraphonic use.



Paraphonic articulatory tongue settings

14.4. As fig 14.4 shows, by reference to the vocoid space in the mouth, in addition to a normal (or laminal) tongue setting, a few others which are possible, and more or less frequent, in languages or accents. Let us carefully notice the deformations undergone by the vowel space, when one of such settings is activated.

fig 14.4. Five frequent paraphonic articulatory tongue settings.



We show these changes especially for the vocoids, because the mechanism is easier to explain, making a constant reference to that restricted area in the mouth space. But, of course, contoids are also involved in each of these settings, according to their articulatory nature.

Other paraphonic elements

14.4. It is important to take into consideration the *rate* of speech tied to particular paraphonic labels, by comparing what can be considered as ‘normal’ to the two logical deviations, ie *slowness* ⟨⟩ as in boredom: *What a bore he is!* and *quickness* ⟨⟩, which is typical of hurriedness: *Quickly now, you two, daddy’s waiting in the car!*

Even the *rhythm* of speech is paraphonically relevant. Therefore, we have to pay attention to whether utterances are said more *rhythmically* than usual, ⟨≈⟩, as for sarcasm: *Oh, you’re very smart indeed!* Besides, *rhythmicity* is still more evident in childish singsongs: *You’re such an idiot!* The opposite element, *a-rhythmicity*, ⟨≠⟩, may occur in shyness: *I’m terribly sorry...could you be so kind as to tell me... where the station is?*

The paraphonic use of phono-articulatory force, of course, is extremely important; most of all *strength* ⟨^u⟩, as in aggressiveness: *I’ll bash your face in!*, but in liveliness as well: *Come on, let’s go dancing!* The opposite is *weakness* ⟨_i⟩, for fondness: *He’s a wonderful person indeed!*

In addition to all this, it may be important to notice whether utterances are fluent or broken up by frequent *hesitation pauses*, which –as we saw at the end of § 13.27– may be ‘silent’ (or unfilled) ⟨|⟩, or ‘voiced’ (and filled) ⟨m̩⟩. The latter are by far the most annoying ones, and may be rendered with other more realistic language-specific symbols: ⟨ʒ, v, ʌ⟩, as for: *Er, well – he’s got er...*, even in phonetic transcriptions: [ʒʒʷEEɪ hɪzˈgɒʃʒʒ].

Even particular types of *articulatory quality* can be added, such as *labialization* ⟨c⟩ in pouting: *No, I won’t go any more!*, or *nasalization* ⟨~⟩ which is possible in boldness: *I won’t speak even if you kill me!*, or the *backing* of the body of the tongue ⟨-⟩, frequent in disgust: *This coffee is undrinkable!*

Other settings, which can be useful for regional descriptions, are: *pharyngealization*, ⟨ˤ⟩ (typical, for instance, of the broad speech of Rome), *faucalization*, ⟨ʌ⟩ (typical of Liverpool English, or again of the broad speech of Catania, in eastern Sicily, cf fig 14.5), and *stiff jaw*, ⟨˥⟩ (of so many English accents), and *stiff lips*, ⟨˦⟩ (again, of some English accents).

Lastly, when we have a *cold*, our voice is chilly with *denasalized* nasal (and nasalized) segments; this is marked by ⟨˧⟩: *I won’t come tonight*. The paraphonic notation is much more convenient than trying to adequately represent the affected segments as in [aə̃˧wɜɔŋ ˈkħɐ̃m ʃə̃˧nəʃ], since we should not ignore (both phonetic and phonemic) vowel denasalization either, as in French. If we do a simple test, we can ascertain that it is nothing like ‘[aə̃˧wɜɔŋ ˈkħɐ̃b ʃə̃˧dʌʃ]’, or in French ‘[ø̃˧bo ˈvɑ̃ ˈblø̃]’, instead of ⟨˧ [ø̃˧bø̃ ˈvɑ̃ ˈblø̃]⟩ *un bon vin blanc*. These articulations are ac-

fig 14.5. Other paraphonic articulatory settings.

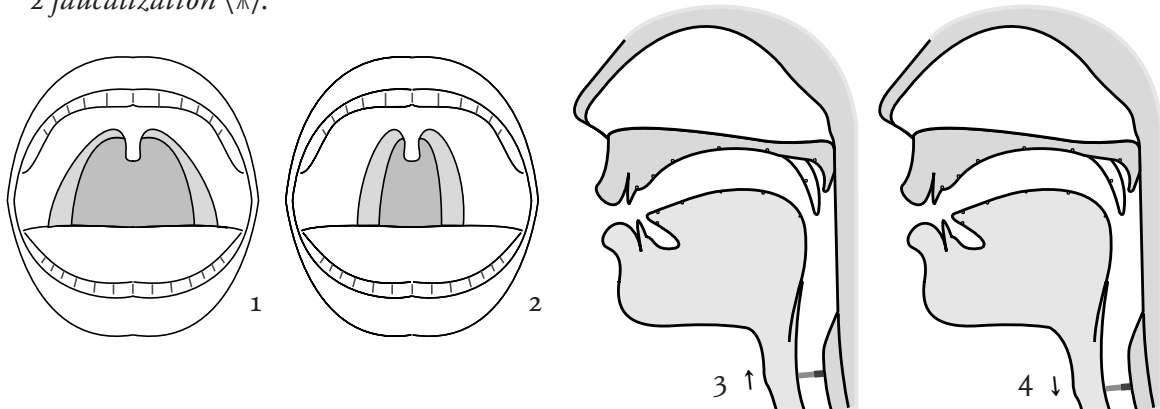
Front view of the open mouth:

1 normal

2 faucalization ⟨ʌ⟩.

3 side view for a *raised larynx* ⟨:·⟩

4 side view for a *lowered larynx* ⟨:·̣⟩.



tually nasal or nasalized, the velum being lowered; but, the air does not manage to pass through the nostrils (either completely or in part), but it goes just to the back part of the nasal cavity, if not slightly further.

Let us note that a *nasal voice* ⟨~⟩ is quite another thing from a denasalized voice (or cold voice, in spite of common belief and terminology). As a matter of fact, a nasal voice can be produced by a certain lowering of the velum, but it can also be produced by some particular resonance in the nasal cavities (with no lowering of the velum), which causes a kind of local vibration. This is typical of a number of accents, also English accents; or it may be an individual characteristic.

14.5. With reference to *phonatory quality*, it is useful to notice whether particular types of voice occur (and to try to detect them). This also goes for articulatory quality, through adequate observations and kinesthetic considerations on one's own articulatory and phonatory movements.

It could be possible to find *murmur voice* ⟨~·⟩ in sadness: *It's a shame it's all over!*, or *whisper voice* ⟨∞⟩ for conspiracy: *We have to act with utmost secrecy!*, or *trembling voice* ⟨?⟩ of old age: *I remember little Joe perfectly well, that rascal!*, or *tense voice* ⟨÷⟩ of worry: *Yes, but what are we going to do?*, or *harsh voice* ⟨×⟩ of fatigue: *I really can't cope any more!* (cf fig 14.2-3).

Besides, *laryngealized voice* (or creaky voice) ⟨ʔ⟩ of sleepiness (seen above *I could sleep like a log!*), or *false voice* (or falsetto) ⟨*⟩ when men put on women's voices: *These high heels are going to kill me!*, or *raised-larynx voice* ⟨:·⟩ as in arrogance: *You obviously don't know who I am!*, or *lowered-larynx voice* ⟨:·̣⟩ as in laziness: *I may do it tomorrow.*

14.6. In addition, the shape and dimensions of Morgagni's ventricle (ie the space between the true and false vocal folds, cf fig 14.6) determines a good part of the individual characteristics of the voice; these characteristics are then accentuated or modified by the supra-laryngeal cavities.

fig 4.6. Vocal folds (or, less well, ‘vocal cords’, 3) and ‘false vocal cords’ (1). The space in between is called Morgagni’s ventricle (2).



A typical *black voice* ($\langle \diamond \rangle$, either African or Black American), makes a particular use of tonality, as fig 14.7 shows. In fact, the three typical intonation bands (of our tonograms) undergo a deformation. Thus, the middle band is expanded vertically, so that the other two bands are, by consequence, restricted. In fig 14.7 this process is indicated as a transformation with ‘re-normalization’.

fig 14.7. ‘Black voice’ paraphonic use of tonality, $\langle \diamond \rangle$.



A classification of individual voices

14.7. Individual voices can be classified according to two parameters: *register* and *resonance*. For any given voice, these parameters respectively refer to an average of their height and vibration, obtained by taking into account the intermediate excursion point of each parameter, according to statistical frequencies of occurrence. They range on a five-degree (though continuous) scale, departing from a middle –un-marked– value; and extending to the cells on the border of the diagram (or *idiophonogram*, marked with ‘+’ or ‘–’).

Further important information about the speakers’ (or, rather, in this particular case, the singers’) sex and age can be supplied in a simple way, as we shall see in a moment. Of course, we must keep in mind that sex and age characteristics form

fig 14.8.A. Idiophonogram for the classification of individual voices.

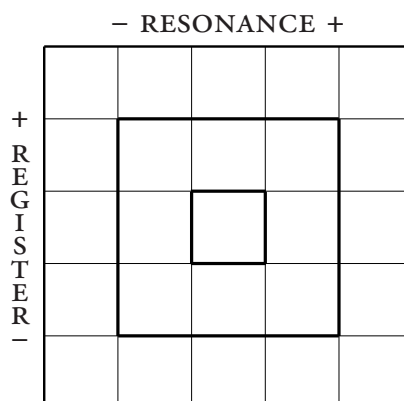
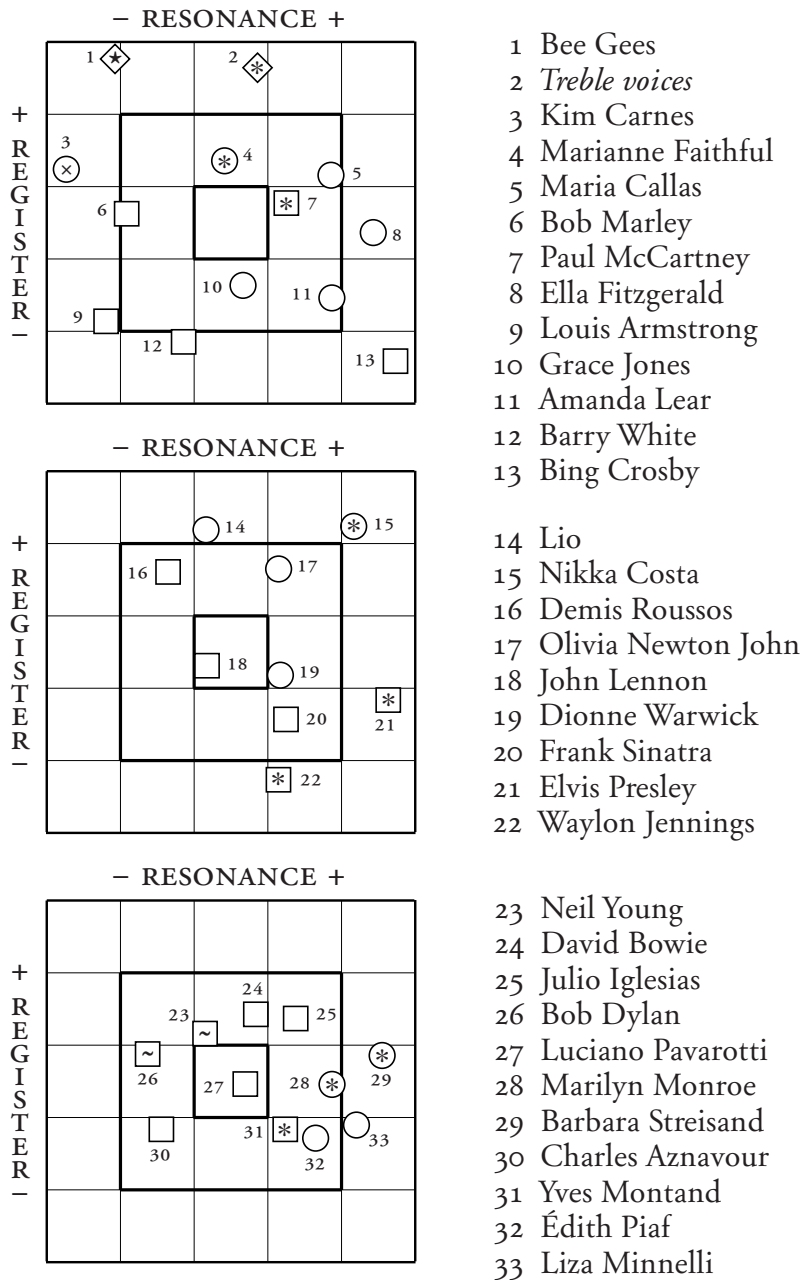


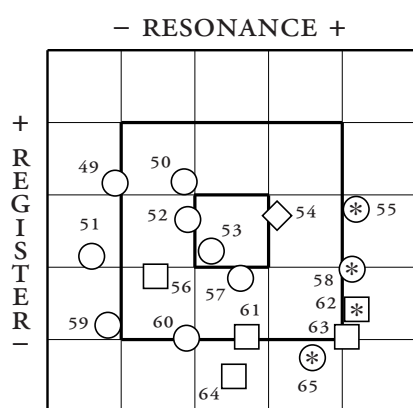
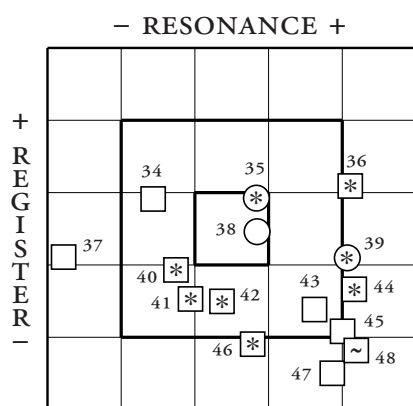
fig 14.8.B. Idiophonograms with 33 voices.



social stereotypes, which depend –at least– on three factors: *physiological* (such as anatomical dimension), *cultural* (such as social and ethnic habits), and *individual* (such as states of health). But, thanks to our own experience, we can generally manage fairly well to make our judgements about the speakers sex and age (except for really ambiguous cases). All this happens in spite of the (sometimes great) differences in the use of paraphonic features made by the various cultures all over the world.

It is not so hard, however, even to determine whether a newborn baby is male or female from its first wailings. As a matter of fact, generally, a newborn-male voice has intense resonance and stronger strength, and a relatively lower register,

fig 14.8.c. Idiophonograms with 32 further voices.



- 34 Little Richard (also with *)
- 35 Astrud Gilberto
- 36 Paul Anka
- 37 Ray Charles
- 38 Mariah Carey
- 39 Patsy Cline
- 40 John Denver
- 41 James Taylor
- 42 B.B. King
- 43 Gene Autry
- 44 Frankie Laine
- 45 Fats Domino
- 46 Don Williams
- 47 Jim Reeves
- 48 Willie Nelson
- 49 Billie Holiday
- 50 Dinah Washington
- 51 Norah Jones
- 52 Chrissie Hynde (*Pretenders*)
- 53 Blondie
- 54 Beach Boys
- 55 Françoise Hardy
- 56 Harry Belafonte
- 57 Judy Garland
- 58 Doris Day
- 59 Amy Winehouse
- 60 Nina Simone
- 61 Dean Martin
- 62 Hank Williams
- 63 John Lee Hooker
- 64 Johnny Cash
- 65 Tracy Chapman

though only a little, and still (very) high, even in comparison with child voices, not to mention (both female and male) adult voices. A newborn female voice, instead, generally has reduced resonance and weaker strength, and a relatively higher register: very high, then.

Sex can be indicated by means of a dot for *female*, or a square for *male*; when no distinction is needed or possible, as for particular groups, we use a rhomb. *Age* can be indicated according to four general classes: *childhood*, *youth*, *adulthood*, and *old age*. They could be shown by attaching a little bar to their symbol of sex, respectively: under it, on its left-hand side, upon it, or on its right-hand side.

Of course, the two five-degree continuous parameters of register and resonance, and the sex and age classifications are only a rough approximation to a full identification of individual voices. As a matter of fact, each voice shows—in different ways—particular mixtures of individual paraphonic characteristics, together with

more or less remarkable idiophonic, social, and geographical features. All of them could be singled out and quantified, and then indicated by means of the appropriate phonetic, prosodic, and paraphonic signs. Of course, the last-mentioned ones are those for pitch, strength, duration, emission, and articulatory and phonatory qualities, plus –at least– two further ones: *brightness*, when a voice (not: pronunciation) is distinct and clear ⟨*⟩; or, on the contrary, *harsh* voice ⟨×⟩, or even ‘falsetto’, or *false* voice ⟨*_{*}⟩; and, lastly, *nasal* voice ⟨~⟩. So, a few voices, although located nearby in the diagram, may still be different enough, because they have –or lack– (more or less frequently) certain paraphonic features.

Some illustrative examples –mostly singing in English– are shown in fig 14.B-C. They have been chosen among (rather) well-known entertainment people, whose actual voices can be heard quite easily on the radio and from recordings practically all over the world, although most of them are no longer performing, or living.

Let us simply add that the paraphonic symbol for *singing* is ⟨♪⟩ (cf § 11.20).

