

Rising-falling contours in speech

A metaphor of tension-resolution schemes in
European musical traditions?

Evidence from regional varieties of Italian

Antonio Romano

Università di Lecce, Italy



1. Introduction

In the literature on music and speech, we are often faced with the idea of a correspondence between the general rising-falling pitch contour in spoken sentences and the classical subdivision of melody in western European music into two parts: a “proposal”, generally ending on a fifth, and a “response”, following a resolutive pattern towards the tonic.

On the one hand, we find a number of references showing how intonative contours can vary dramatically across languages; nevertheless, in most cases they tend to appear in accordance with a “rising-falling pattern”. On the other hand, as recently shown by the music theorist Eugene Narmour (1990) and further discussed by Russo & Cuddy (1999), there are patterns of melodic motion that appear to transcend many musical styles and that seem to be an extension of common vocal melodic patterns.

Local contours as well as overall contours appear to differentiate question sentences from statements: interrogative patterns in many languages tend to show a final rising of the vocal pitch, while affirmative clauses tend to “precipitate” on the tonic. Now, dialectal varieties of Italian, as well as Romance dialects scattered throughout the Peninsula, may present different intonative contours, and yet, something like a relative “fifth jump” can be almost always recognised in the *catastasis* of question sentences where a local contour marks a melodic movement rising different kind of expectation.

The data discussed in this contribution derive from field enquiries held in Aosta Valley, Apulia Region, Südtirol and Abruzzi Region. Other results are

provided by a pilot study carried out on the perception of melodic contours of sentences from different western European languages. They show the presence of (a) a suspension in the melodic contour of sentences according to the theory of “suspended meaning”, valid both in music and language; (b) native speakers of different varieties recognise these patterns just listening to partial intonation contour of synthetic stimuli.

Rise and fall in music

A classic subdivision of melody in western European music tradition usually distinguishes two parts in melodic phrasing: a first part, called “proposal” (or “question”), generally ending on a fifth, and a second part, called “response” (or “answer”), following a resolute pattern that tends to globally fall towards the tonic.

It is a common experience to listen to structuring melodies with increasing expectations in the first part, finally resolved with a final concluding tune. A brief passage of D. R. Hofstadter clearly resumes this kind of experience:

“ [listening to music] we maintain a mental stack of keys, and [...] each new modulation pushes a new key onto the stack [...]. Any reasonably musical person automatically maintains a shallow stack with two keys. In that “short stack”, the true tonic key is held, and also the most immediate “pseudotonic” (the key the composer is pretending to be in). In other words, the most global key and the most local key. That way, the listener knows when the true tonic is regained, and feels a strong sense of “relief”. The listener can also distinguish [...] between a local easing of tension — for example a resolution into the pseudotonic — and a global resolution. ” (Hofstadter 1979: 129).

This sense of fulfilment or finality in falling pitch is also described by Cooke (1959: 104) whereas in Meyer (1956) we find accounts for a distinct consideration of such a sensation related to an affective response but dominated by a conscious expectation.

“ Whether a piece of music gives rise to affective experience or to intellectual experience depends upon the disposition and training of the listener. [...] Thus while the trained musician consciously waits for the expected resolution of a dominant seventh chord the untrained, but practiced, listener feels the delay as affect ” (Meyer 1956: 40).

As a consequence of affect or rationality, we could see that, in any way, expectation in music involves a high order of mental activity. The fulfilment of a habit response, in art as well in daily life, requires judgement and cognition

both of the stimulus itself and of the situation in which it acts.

Thus, an immediate correspondence of this sensation has to be searched for in the structure of the stimuli themselves but also in the way how we perceive the melodic stimulus within the frame of our experience and our personal taste.

A general metaphor linking pitch and suspense could be seen in the fact that pitch is felt by everyone to be an 'up-and-down' dimension. It could be claimed that there is no reason for calling notes with more vibrations per second 'higher', except in so far as they have always been written higher on staff. In answer to this, Cooke (1959) points out the connections between the following facts: (1) by the law of gravity, 'up' is an effort for man, 'down' a relaxation; (2) to sing 'high' notes, or play them on wind, brass, or string instruments, demands a considerable effort; (3) to tune a string 'upwards', one screws 'up' its tension; (4) scientists, talking of 'high' notes, speak of a 'high' number of vibrations per second.

Another possible explanation is given by I. Fónagy: "*Cette projection spatiale du ton est justifiée par le fait qu'il est plus facile de produire une note élevée en levant la tête, et une note plus basse en baissant le menton*" (Fónagy 1983: 121).

Rise and fall in speech

In speech we observe very common patterns of suspension which leave interrogative sentences (and, to some extent, also declarative not concluded) open towards a possible way of integration — expected by the listener — or introducing a possible continuation in the linguistic program of the speaker. These suspensive patterns succeed in the generation of a tension by means of a tonal rise. Only a declining contour allows the satisfaction of this strong feeling by producing the resolution of the mental tension.

An extended sample of languages can be described as having a globally rising-falling pitch movement in the single intonation unit of a simple unemphatic declarative utterance — where the overall pattern generally finishes on an extreme low pitch.

Exceptions to the general rule are mentioned for dialect variants as in some Midland and Northern dialects of British English, as well as in Estremadura dialect of Spanish and in the Corfou dialect of Greek, where declaratives are said to end with a rised final pitch (Hirst & Di Cristo 1998: 19). But we would easily add in this list, without hesitation, Venetian dialects and north-eastern

regional varieties of Italian, whose declarative sentences are often perceived as questions by other Italian people.

2. Downtrends and uptrends

Physiological explanations in terms of universal constraints have been proposed for the declination of F_0 within affirmative sentences in a number of languages.¹

The movement in the fundamental contour, accounting for accent components and sentence mode components, would be globally characterised by a downtrend whose mean slope is defined as a consequence of the programmed sentence length and of the number of *downsteps* or *upsteps* marking accent realisation. At this purpose one can observe how a declination of smaller units is reset into declination of larger units (*declination within declination*, also see Ladd 1996).

The overall *downward sloping* is described as a relaxation gesture in Lindblom (1968) and as a principle of articulatory laziness in Ohala (1990). In Vaissière (1983) we found that the overall tendency to the declination may be disabled, as it happens in question sentences. There the trend is reversed, and an uptrend appears at the end of the sentences. Lieberman (1967) studies the final rising of questions within breath groups and describes it as a product of an increased activity of larynx muscles, but several other approaches attempted to give full explanations of this phenomenon.

Independently from the original reasons of such overall trends, we can model them from the functional point of view, by analysing each typical prosodic configuration as a result of the recursive implementation of similar patterns. Therefore, the basic rise-fall pattern would be determined on biological bases and as an effect of the phonologisation of the archetype "conventionalised" as the phonetic marker of the completeness of an utterance in many languages. Instead, in this framework, all the non-fall patterns would have psychological and ethological explanations as well as in the phonologisation of the contrast (see Vaissière 1995). J. Ohala defines an "ethological frequency code" and observes that "low frequencies signal domination: so a person making a statements uses a low frequency. High frequencies signal submissiveness. A person asking a question [...] tends to use a high pitch voice." (Ohala 1984). Moreover, as it is clearly discussed by Vaissière (1995), regardless of size, each constituent tend to conform the shape of a common archetypal (rise-fall)

contour and a few derived contrastive (rise-non fall contour, whose characteristics may be motivated on biological, psychological and/or ethological grounds).

Declination in Italian

Classic reviews on the prosodic characterisation of standard Italian describe a final declination for assertive utterances and a rising movement (global or at least at the end of the sentence) for yes/no-questions, regional varieties seem to have recourse instead to alternative patterns to differentiate questions from statements in spontaneous speech.

Instrumental evidence of rising-falling contours signalling questions comes from various researches carried out by various authors and accounting for Sicilian Italian of Palermo, southernmost Sallentinian, different varieties of Sardinian, Apulian Italian of Bari as well as for some francoprovençal varieties of the Aosta Valley. The auditive impression deriving from the observation of patterns in Neapolitan or, within a different rhythmic framework, in some Italian varieties spoken in Piedmont or even in Tuscany, suggests the existence of more varieties in which this principle of distinctiveness needs more accurate study. Further studies should be also devoted to the intonative system of Venetian (showing a declarative scheme always tending to rise at the end) in order to investigate the way how it has been redesigned to maintain the statement-question opposition.

As a by-product of a fieldwork in some apparently homogeneous Italian linguistic areas, we detected different ways to signal question in terms of specific local contours — as geographical distinctive features — but we observed the same underlying melodic strategy: creating a tension and resolving it. Yet, the communication between people coming from areas where different prosodic sub-systems are widespread is not prevented by these cues.

Therefore we tried to observe how the rising is really implemented and how people behave in front of musical stimuli presenting these different situations.

The real implementation of a rising to signal interrogation

We considered sentences from two Italian speakers marked by a different use of melodic contour (but with the same rhythmic properties). As we showed in previous publications, though sharing the same overall linguistic system, Sallentinian speakers, have recourse to prosodic systems mainly differing on

the final contour of *y/n* question which is rising-falling for speakers of the southernmost area (as the male speaker FC29 here considered, see Figure 1), and always rising for speakers of the central-northern area (as the female speaker FM28, see Romano, 1997).

Both contours in 1a. and 2a. are characterised by a rising-falling movement ending towards a low pitch (-1 tones compared to the mean pitch values just before the last stressed vowel).

A final rise is realised upon the last stressed syllable in 1i. ($+2$ tones) followed by a fall-down on the post-stressed ones ($-1 \div +2$ tones) vs. a final global rise in 2i. mainly involving the final unstressed syllables (about $+1$ tone).

As we can resume, apparently no fifth jump seems to be regularly realised: but for speaker 1 two different movements are opposed upon the last stressed syllable generating a gap of at least 3 tones; for speaker 2 a gap becomes sensible only upon the post-stressed phase and reaches its maximum at the end of the sentence (hence only about $+2$ tones, even if this gap has been observed spanning up to $+4$ tones in more emphatic sentences).

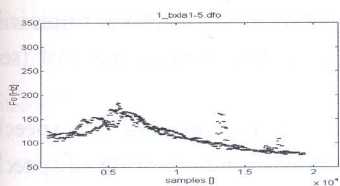
So, the claim that standard Italian has a rise to a fifth, in common with musical works, can only be a statistic reference in view of the wide differences in pitch range found between speakers on the one hand, and between utterances of one person in different states of mind and different communicative situations.²

Nevertheless, as different contours are used by speakers geolinguistically distant to realise the "same" message, within their own intonative codes, slightly different aesthetic preferences appear in the perception of "foreign" prosodic features and both scientific and popular terminology on these patterns refer to a musical setting and to singing.

As highlighted by I. Fónagy, evident analogies do exist between melodic patterns in emotional speech and traditional schemes in European music. Intonation brings us towards the separation phase between speech and music, to an hypothetical ancestral language, whose unique function was the resolution of biological and mental tensions (Fónagy 1983: 149).

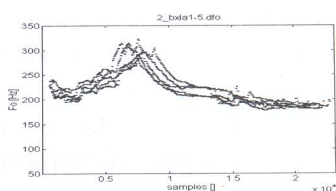
We claim that, independently from the selection of a different final contour, relevant acoustic cues of an interrogative characterisation of the sentence are given by the realisation of an unresolved tension in the first part of the utterance and by a functionally distinctive way to oppose a resolute asserting pattern to a suspending one: at any rate the tension generated at the end of a question contains marks signalling that the question is achieved.³ Our claim seems to be supported by the results of the perceptual test described below.

1a.



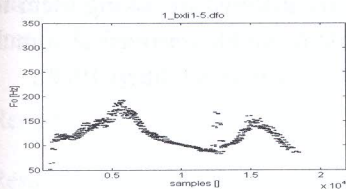
[labambi:navwɔlelabam b ola .]

2a.



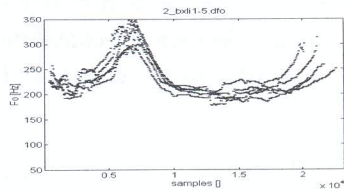
[la bambi:navwɔlelab am b ola .]

1i.



[labambi:navwɔlelabambola ?]

2i.



[labambi:navwɔlelab ambol a ?]

Figure 1. Pitch contours of 5 realisations of 2 sentences (declarative and interrogative, corresponding to the same order and organisation of speech units) uttered by two Italian speakers (1a. and 1i. by a 29 years old, male speaker of a southern Sallentinian variety; 2a. and 2i. by a 28 years old female speaker of a northern Sallentinian variety). The sentence in 1a. and 2a. is *La bambina vuole la bambola.*, “The child wants the doll”. The same sentence is realised in two distinct ways in 1i. and 2i. at the interrogative mode, with the same meaning nuance and the same information distribution (narrow focus on the last constituent). The question is *La bambina vuole la bambola?*, “The child, she wants the doll? (or something else?)”.

3. A perceptual experiment

As for the principle of pattern perception, L. Meyer said that

“To assert that incompleteness gives rise to expectations of completeness is tantamount to tautology. For things seem to be incomplete only because we entertain definite feelings, latent expectations, as to what constitutes completeness in a given stimulus situation” (Meyer 1956: 128).

Our sense of completeness or incompleteness is also a product of those patterns or sound terms which become established as more or less fixed, given parts of a particular work. That is, once a sound term has been established as a coherent, though not necessarily as a complete or closed unit, then part of the series taken by itself will, generally speaking, seem to be incomplete, particu-

larly if the fragment occurs in the earlier parts of the total work. Thus repetitions of the beginning of a well-shaped pattern already heard several times will arouse expectations that it will be completed as it has been in the past (see Meyer 1956: 128–129).

The perception of different ways to realise such an expectation in speech has been investigated by trying to evaluate the rate of identification of speech contour segments.

We carried out the experiment with listeners who had an active competence of one of the two intonative sub-systems analysed by asking them to express a judgement of completeness and sentence mode on musical stimuli obtained from such segments of both varieties (cf. Ohala & Gilbert 1979).

Stimuli

We selected three declarative-interrogative pairs of utterances as the ones reported in Figure 1, and we prepared a series of 12 stimuli randomly organised to be used in the perceptual testing of one sample of listeners. The task of identifying the completeness of the sentence being easy enough in the case of natural utterances containing verbal information, only the melodic information was taken into account. The stimuli were sequences of pulse trains generated on the base of the rhythmic and melodic organisation of the contours extracted by an electronic pitch analyser and used to generate synthetic stimuli by means of software developed at the purpose. The original sentences were one declarative-interrogative pair pronounced by the two speakers (4 utterances) and another pair of longer sentences uttered by only one speaker (the southern one). The sentences were cut into two pieces after the point where the highest tension was reached. Then the test sequence was made by 6 initial and 6 final portions randomly ordered.

Procedure

Every subject listened to the random sequence of stimuli in isolation with headphones and was free to hear the stimulus whenever he/she wanted by clicking on a computer screen. The stimuli, corresponding to each of the 12 sentence portions of the corpus, were presented to the listener in random order asking him/her to freely rate them by clicking on one of 4 buttons proposing an answer such as “1st part of a declarative sentence”, “2nd part of a declarative sentence”, “1st part of an interrogative sentence” or “2nd part of an interroga-

tive sentence". A first test, with free listening of the stimuli, was aimed at training the listener to the test conditions.

Subjects

Twelve subjects were drawn from various places of the region and included undergraduate and graduate students, school teachers and local travellers which had or not experience of different varieties. They were organised into two groups after their original geolinguistic competence: 6 listeners were from the southern area (sub-system 1.) and 6 from the northern one (sub-system 2.).

Results

Results are resumed in the tables below organising data in confusion matrices.

As a general trend, none or only a few of the non-final segments of a sentence are rated as final question contours and this for both groups. First parts of a question can be confused with first parts of an assertion and vice versa, but they are rarely misperceived as final parts (resolutive or not).

The two groups of listeners did not show significant differences in the way they distinguish a question from a statement or they rise expectations or fulfil them listening to different stimuli marked by slight geolinguistic diversity. Segments of a declarative sentence have been correctly identified in 75% of stimuli. The same score has been reached for interrogative stimuli. A suspensive initial contour seems to have been perceived in 85% of the cases in which the stimulus was really incomplete but initial. For stimuli coming from the last incomplete part of a question the identification rate has been slightly lower (81%). These results do not support a pretended innate experience of variable tension-resolution schemes — a passive competence of "foreign" patterns could give relevant cues at this purpose. Such an evidence may come instead in other ways. A pilot experiment is currently being carried out on several European languages with a group of subjects having no experience about them, as opposed to a group of subjects who learned or who were studying them. The two groups did not show significant differences in the way they distinguish a question from a statement in a foreign language on the basis of their melodic contour alone.

Table 1. All the stimuli (declarative and interrogative sentences) and both groups.

Rated as	Stimuli			
	Statement		Question	
	1st part	2nd part	1st part	2nd part
1st part of an incomplete statement	47%	11%	36%	6%
2nd part of an incomplete statement	28%	64%	3%	6%
1st part of an incomplete question	22%	8%	64%	6%
2nd part of an incomplete question	6%	14%	3%	78%

Table 2. First part vs. Second part of a sentence

Rated as	Stimuli	
	1st part	2nd part
1st part	85%	15%
2nd part	19%	81%

Table 3. Statement vs. Question

Rated as	Stimuli	
	Part of a statement	Part of a question
Part of a statement	75%	25%
Part of a question	25%	75%

4. Conclusions

In this paper we resumed some elements of the correspondence between general rising-falling contours in speech and the classical European musical schemes proposal-response.

As already stated in a number of different works, intonative contours of sentences tend to appear, across languages, cultures and dialects, in accordance with a rising-falling expectation pattern. As a matter of fact, in questions the "falling" has generally to be seen as an actual overall rising in pitch differently shaped from one variety to another.

We analysed dialectal varieties of Italian, as well as Romance dialects spoken in Italy, which may present slightly different intonative contours in

question sentences, and yet, a relative "fifth jump" has been often recognised in the *catastasis* of the sentence while a local contour appeared at the end of the sentence marking a complete melodic movement in terms of the listener expectations but referring to an incomplete knowledge of the speaker that needs for a response.

In several field enquiries we observed that people tend to associate emotional or stylistic or geolinguistic labels to different patterns, but in most cases they correctly identify a broken contour from a final rising or rising-falling interrogative contour, probably thanks to something like a feeling of fulfilment, exactly like if it were music.

We claimed, and outcomes from our experiments partially proved, that a common experience of rising-falling pattern as well as of tension-resolution schemes provides solid cues to the listeners in view of the identification of semantic completeness. Our first findings agree, under certain conditions, with the conclusions of Meyer who guessed "expectations based upon learning" prior to the natural modes of thought. The expectations which are entertained on these basis are actually associated to structural gaps to be filled; but what constitutes such a gap depends upon what constitutes completeness within a particular system (Meyer 1956: 43). At this stage, it becomes uncertain to consider rising-falling contours in speech as a metaphor of tension-resolution schemes in singing or to see both of these schemes on a common level of structural derivation from more basic (universal), possibly semiotic, principles.⁴

A common semiotic function might be a product of signature properties of more primitive motoric or limbically controlled events. A parallel may be seen in the temporal structuring of music and speech, where phrase final lengthening, while undoubtedly having an important signalling function, is a product of all motor controlled sequencing activities (as it has been suggested by our anonymous reviewer).

Notes

1. This phenomenon is investigated in the literature, including the research of Lieberman (1967), and the models developed by Gårding (1979). Further details about these topics are also in Bruce (1982), Bertinetto & Magno-Caldognetto (1993) and Ladd (1996).
2. Nevertheless, other structural elements across languages contribute to reinforce this model: listings, for example, have recourse to melodic elements intended to generate a

strong "suspended meaning" just before the last item of the list, the penultimate item being realised with an overall upward and the last one within the frame of a downward.

3. At this stage, two types of incompleteness can be distinguished: one which arises in the course of the pattern because something was skipped over; another one in which the figure, though complete so far as it goes, simply is not felt to have reached a satisfactory conclusion, is not finished. The first type of incompleteness may be seen as a "structural gap", the second type, as a product of a delay in the need and desire for "closure". And if, for the latter we can assume a linguistic learned competence, for the former, we suspect a "competence" related to music and to the experience of natural laws.

4. The emergence of singing has been related to melodies with a given meaning, a power to communicate but also, maybe, a distinctive power for the personality of the singer or speaker: a mark of his/her originality (Collaer, 1965: 625). The origin of mimesis, language gestures and human expressions, as a communication medium, are probably related to an evolution stage preceding the emergence of the articulated languages. Like in some *apothroptic* songs, these gestures may respond to the idea of the imposition of human will (cf. the *respite from death gained by a tale* in Abry, 1997).

References

- Abry, Christian (1997). Pour une Histoire Naturelle de la Parole dans la "Théorie de l'Esprit". *Thèse d'habilitation*, Univ. de Grenoble (France).
- Bertinetto, Pier Marco & Emanuela Magno Caldognetto (1993). "Ritmo e intonazione". In A. A. Sobrero (Eds.), *Introduzione all'italiano contemporaneo. Le strutture*. Bari, Laterza, 141–192.
- Bruce, Gösta (1982). Developing the Swedish intonation model. *Working Papers in Linguistics*, Univ. Lund, 22, 51–114.
- Collaer, Paul (1965). "Il mondo della musica". In V. L. Grottanelli (Ed.), *Ethnologica. L'uomo e la civiltà*. Milano, Labor.
- Cooke, Deryck (1959). *The Language of Music*. Oxford University Press, London & New York.
- Fónagy, Ivan (1983). *La vive voix*. Paris, Payot.
- Gårding, Eva (1979). Sentence Intonation in Swedish. *Phonetics*, 36, 207–215.
- Hirst, Daniel & Albert Di Cristo (1998). "A survey of intonation systems". In D. J. Hirst & A. Di Cristo (Eds.), *Intonation Systems: a Survey of Twenty Languages*. Cambridge Univ. Press, 1–40.
- Hofstadter, Douglas R. (1979). *Gödel, Escher, Bach: an Eternal Golden Braid*. New York, Basic Books.
- Ladd, D. Robert (1996). *Intonational phonology*, Cambridge, Cambridge Univ. Press.
- Lindblom, Bjorn (1968). Temporal organisation of syllable production. *Quarterly Progress Status Report of the Speech Transmission Laboratory*, Stockholm, 2, 1–15.
- Lieberman, Philip (1967). *Intonation, Perception, and Language*. Research Monograph No. 38, Cambridge, Mass, MIT Press.

- Meyer, Leonard B. (1956). *Emotion and Meaning in Music*. Chicago, University of Chicago Press.
- Narmour, Eugene (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago, University of Chicago Press.
- Ohala, John (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, 41, 1–16.
- Ohala, John & J. B. Gilbert (1979). Listeners ability to identify languages by their prosody. *Studia Phonetica*, 18, Ottawa, Didier, 123–131.
- Romano, Antonio (1997). Persistence of prosodic features between dialectal and standard Italian utterances in six sub-varieties of a region of Southern Italy (Salento): first assessments of the results of a recognition test and an instrumental analysis. *Proc. of EuroSpeech'97*, Rhodes 1997, 175–178.
- Russo, Frank A. & Lola L. Cuddy (1999). Motor Theory of Melodic Expectancy. *Proc. of the Acoustical Society of America — ASA/EAA/DAGA '99 Meeting* (Berlin, 1999).
- Vaissière, Jacqueline (1983). Language Independent Prosodic Feature. In A. Cutler & D. R. Ladd (Eds), *Prosody: Models and Measurements*, Berlin, Springer, 53–66.
- Vaissière, Jacqueline (1995). Natural Explanations for prosodic cross-languages similarities. *Proc. of ICPhS 95*, (Stockholm, 1995) 654–657.