HAL
open science

RISE DYNAMICS DETERMINES TUNE PERCEPTION IN FRENCH: THE CASE OF QUESTIONS AND CONTINUATIONS

Lydia Dorokhova, Mariapaola d'Imperio

To cite this version:

Lydia Dorokhova, Mariapaola d'Imperio. RISE DYNAMICS DETERMINES TUNE PERCEPTION IN FRENCH: THE CASE OF QUESTIONS AND CONTINUATIONS. International Congres of Phonetic Science ICPhS2019, Aug 2019, Melbourne, Australia. hal-02476914

HAL Id: hal-02476914

https://hal.science/hal-02476914v1

Submitted on 13 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RISE DYNAMICS DETERMINES TUNE PERCEPTION IN FRENCH: THE CASE OF QUESTIONS AND CONTINUATIONS

Lydia Dorokhova[1] & Mariapaola D'Imperio[1,2]

[1]Aix Marseille University, CNRS, LPL UMR 7309, Aix en Provence, France
[2]Rutgers University, New Brunswick (NJ), USA
lydia.dorokhova@lpl-aix.fr, mariapaola.dimperio@lpl-aix.fr

ABSTRACT

The present study aims at uncovering the impact of dynamic details of intonational rises sharing the same tonal description but differing in the actual shape of the F0 contour. Two French rising final contours, both labelled as LH*H% in AM, are used to mark continuations or yes/no questions. The hypothesis tested here is that a perceptual difference between the two contours would stem from the particular dynamic characteristics of the rise, notably from the shape of the interpolation between the L and the H tonal targets, which has been noticed to be different between the two (convex vs. concave). We hence created resynthesized speech stimuli, in which tonal target alignment, scaling and segmental duration were rendered ambiguous, while rise shape was modified from concave to convex in two base stimuli. Results of a two-alternative forced choice test suggest that dynamic properties of the pitch rise can alone affect contour identification.

Keywords: intonation, perception, interpolation shape, question, continuation, French.

1. INTRODUCTION

Intonation contours have either been modelled in holistic ways, through rises and falls [7, 24], or else as a sequence of static tonal targets [26]. Static target models are based on the notion of 'turning point' within the F0 curve, defined on both time and pitch dimensions (e.g. L(ow) and H(igh) tones). This is because the AM model of intonation stipulates that the shape of the interpolation as well as the slope between two tonal targets is not relevant, since it is the result of a mostly linear interpolation between them [26, 2, 3]. In other words, a rise, for instance, can be steep or shallow, but this would only be an epiphenomenon of the longer or shorter segmental sequence intervening between the targets. Moreover, shape differences in terms of curvature are not taken into account. This type of static description of prosodic events in production is not disputed anymore, though the relevance of shape and slope differences has yet to be dismissed in the realm of intonation perception.

Work reported in [17, 18, 19, 20, 9, 16, 8, 4] has actually shown that complex dynamic phenomena must be taken into account in the process of identifying intonational form. Specifically, dynamic properties of the F0 contour, such as peak shape and slope [8, 21] have been found to be perceptually relevant for pitch accent category assignment. The Tonal Center of Gravity [ToCG, 4] hypothesis, inspired by empirical data on intonation perception in languages such as Italian [8], English [21] and German [22], is based on the idea that perceived properties of tonal contours (both alignment and scaling) are influenced by the integration of their dynamic properties, such as the area covered by a rise-fall shape.

Rising curvature has been found to be nearly significant in the perception of question vs contrastive topic statements in Neapolitan Italian [10, 25], and appears to be difficult to imitate in L2 production, even for very proficient speakers [14]. Moreover, it is differently used by individual speakers of the same variety [23]. Hence in this study we explore whether rising contours, sharing an identical AM description could be differently identified on the basis of curvature differences alone.

In French, both terminal continuation and yes/no question contours can be identically labelled as a LH* H% sequence [27, 12, 31, 28] though their curvature can be different in production. In informal observations, Delattre [13] described the continuation contour as being more concave than the question contour (in addition to having a lower final F0 peak). The perception data in [15] showed indeed an impact of contour shape in perception, though in the opposite direction, i.e. questions were mainly cued by concave rise shape. The hypothesis we explore here is hence that French continuation and question rises can be identified in perception on the basis of rise shape alone, independent of L and H alignment and scaling. A perceptual test with resynthesized stimuli was therefore performed to test our hypothesis.

2. METHOD

2.1. Stimuli

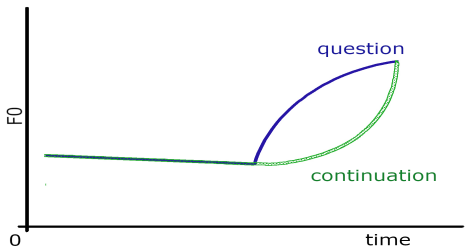
20 utterances composed of equal segmental material were recorded by a female speaker of Southern French with either a question or a continuation contour (20*2). The rise was always followed by an IP boundary, as in the examples in (1) and it always occurred on the

second syllable of the noun preceded by the preposition des 'of the'. To obtain a sentence internal IP boundary, declarative sentences were enriched with a parenthetical after *canaux* 'canals' (see 1a).

(1a) *La beauté des canaux, telle qu'elle est décrite par les touristes, donne envie de voyager.*
'The beauty of the canals, as it is described by tourists, gives a desire to travel'

(1b) *La beauté des canaux ?*
'The beauty of the canals?'

Figure 1: Schematised pitch contour in the two modalities.



Recordings were made with a head-microphone in an anechoic room with Audacity software, at the Parole and Language Laboratory of Aix Marseille University. Utterances of the type shown in 1a above where cut at the start of the parenthetical before creating the resynthesized stimuli, in order to have base stimuli of the same size.

2.1.1. Norming

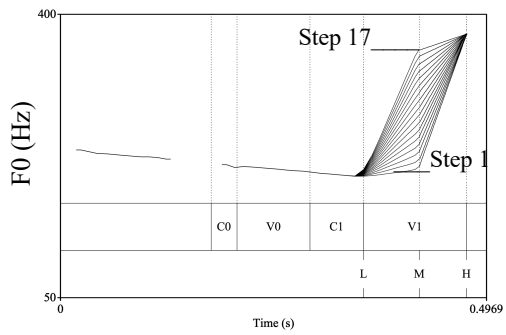
An online survey with a two-alternative forced choice (question vs. continuation) task was carried out on the gorilla.sc website. 9 participants chose a prototypical question/continuation pair among the stimuli. '
Out of the 40 heard utterances only 1 utterance pair received 100% identification for each member, which was then chosen as the basis for resynthesis. 16 additional correctly identified stimuli (over 80% correct) were further used as fillers.

2.1.2. Perception stimuli

Two base stimuli, a prototypical declarative question and a prototypical continuation, were hence selected. From each utterance we then cut the sequence *des canaux* [dekaˈno] ('of the canals'), to which the final rise was associated. Segmental durations as well as tonal target scaling and alignment were first rendered ambiguous by taking the average values of the two original base stimuli. Specifically, for alignment, L targets were placed at the boundary between the last onset consonant /n/ and the final vowel, while H alignment was kept at the offset of the final vowel (see Fig. 2). Tonal targets in both utterances were scaled to the same F0 level (199 Hz for L, 379 Hz for H), i.e. the average between the values of the two prototypical utterances. The initial F0 values in the stretch preceding the L target were also averaged between the two bases and resynthesized as a straight line.

After stylizing the two base stimuli in PRAAT [6], the interpolation line between L and H was modified so as to create two continua going from the most concave (step 17) to the most convex (step 1) shape. In order to do so, the mid-point of the interpolation between L and H was selected (M) and the corresponding F0 value was modified in 10 Hz steps from a value equal to that of the preceding L target (209 Hz) up to a maximum value of 359 Hz (i.e., 10 Hz lower than the value for the H target), for a total of 17 steps (see Fig 2). The two base stimuli (step 0), resynthesized with ambiguous duration but retaining original F0 values, were included in the experiment.

Figure 2: Schema of the 17-step modification with segmental annotations (C for consonant, V for vowel), used to create the 2 continua.



2.2. Participants

30 French native speakers drawn from the university population at Aix Marseille University participated in the study (6 males) with an average age of 22.43 (min. = 18, max. = 40). All participants were monolinguals and did not report any hearing impairment. Participants received monetary compensation in the form of a gift-voucher.

2.3. Procedure

The experiment took place in small groups of 2 to 4 participants in a computer room at the *Parole and Language* laboratory in Aix-en-Provence, France. A two-alternative forced choice task was created through Perceval software [1]. Each participant was placed in front of a computer screen, equipped with headphones and response pads (placed under the dominant hand). Sound intensity was always set to the same value.

Participants were asked to give one of two possible responses upon hearing each stimulus: question or

continuation. They were instructed to give their responses as quickly as possible and could only listen to each stimulus once. After a short training session, experimental stimuli were presented in 5 repetition blocs of 36 stimuli each, plus 16 fillers. A short break was placed after the third bloc to allow participants to rest.

2.4. Results

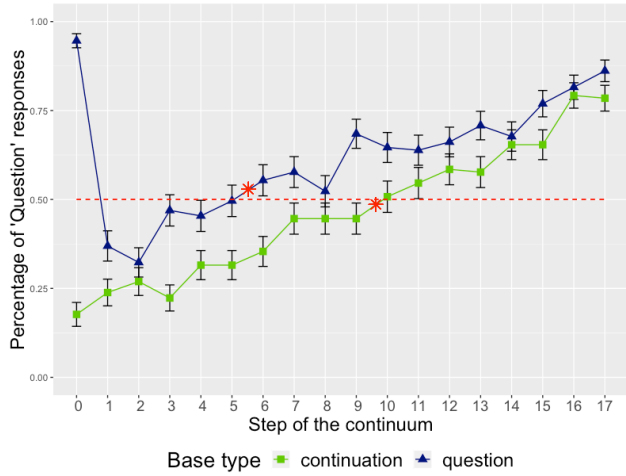
Response scores and reaction times were collected (though reaction times are not analysed here). Data for 4 participants were excluded (they responded with the same key more than 90% of the times). Thus, we analysed data from a sample of 26 participants, for a total of 4680 data points. Percentage of "question" responses was calculated across all subjects, by step (1 for the most convex, 17 for the most concave, as shown in Fig. 2, while 0 is used for the base stimuli).

Note that the base stimuli (step 0) were identified with a very high score (95% of "question" responses for the question-base stimulus and 83% of "continuation" responses for the continuation-base stimulus), despite their duration was rendered ambiguous between the two modalities. Note also that the Step manipulation induced the same direction of the effect for both continua, in that increasing concavity lead to more question responses from stimulus 1 to stimulus 17 for both series.

To statistically test the response variance, mixed-effect logistic regression models were carried out. R software with 'lme4', 'lmerTest', 'car' and 'multcomp' packages was used to carry out the statistical analysis. Step (0 to 17), Base type (question vs. continuation) and Bloc number were specified as fixed factors (with their interaction), and random slopes for Base type and Step as well as random intercepts for Participants, were included in the retained model[1]. Since our data set was too large for the default fitting methods in 'glmer', it was necessary to implement the function option "nAGQ=0" for the model-fitting process to converge [5]. The Anova function retained our three fixed factors Step ($\chi2$=50.75, p<0.001), Base type ($\chi2$=42.44, p<0.001) and Bloc ($\chi2$=6.35, p<0.01) as significant.

As for the interactions, the model retained only the one between Step and Base ($\chi2$=37.21, p<0.01) and Base by Bloc ($\chi2$=6.45, p<0.01), so that multiple comparisons (Tukey test with Bonferroni correction) were carried out to find the source of these interactions. The interaction between Step and Type could be explained by a smaller rate of change in the amount of question responses across step in question-based stimuli than in continuation-based stimuli. This can also be observed both in Fig. 3 and in Tab. 1. The model was then retransformed for multiple comparisons to have 36 levels (18 Steps * 2 Base types), 10 levels (5 Blocs * 2 Base types). For the sake of space, not all comparisons are presented. Table 1 shows the comparisons of 3 steps at the beginning of the continuum (before step 10) and 3 steps after step 10.

Figure 3: Percentage and standard error of 'question' responses by Step and Base for all subjects.



For instance, response score was significantly different between the continuation-base steps 3 and 14, which was not the case for questions.

Detailed analyses of multiple comparisons of the interaction between Bloc and Base type revealed an effect of Bloc 2 (β =-5.06, β SE= 1.428e-01, p<0.01), i.e. of Bloc 3 (β=-5.57, β SE= 1.428e-01, p<0.01) and Bloc 5 (β =-5.88, β SE= 1.431e-01, p<0.01) for continuation responses when compared to Bloc 1. No such dependency was found for questions. Therefore, a slight effect of adaptation to the task was observed only for continuation responses.

Table 1: p values of several multiple comparisons between steps. Significant p values are represented on the darker background.

	Continuation base		
	Step 1	**Step 3**	**Step 7**
Step 10	0.000772	0.000181	1.000000
Step 14	4.39e-10	7.02e-11	0.133679
Step 17	< 2e-16	< 2e-16	9.19e-07
	Question base		
Step 10	0.000884	0.965744	1.000000
Step 14	2.64e-05	0.122697	1.000000
Step 17	< 2e-16	1.38e-09	3.35e-05

We also estimated 50% crossover points for both continua on the basis of the simpler glmer model[2] and

[1] R function: glmer (response~step*type*bloc +(0+step2|subject)+(0+type|subject)+(1|subject), family="binomial", nAGQ=0).

[2] R function: glmer (response~step + (1|subject) family="binomial", nAGQ=0))

obtained the following values: for question base stimuli, crossover occurred at step 5.53 (SE= 1.49), while for continuation base stimuli at step 9.63 (SE=1.27). Step 0 was not included for these calculations. The estimated crossover points for each continuum are represented in Fig. 2 with red stars.

3. DISCUSSION

In this study we tested the effect of interpolation shape (continuation vs. declarative question) in French. Resynthesized stimuli were used in a two-alternative forced choice task with two 17-step continua created on the basis of either a continuation or a question base utterance in which segmental duration was rendered ambiguous between the two modalities.

First, our results strongly suggest that the shape of the LH interpolation can alone shift contour identification from one modality to the other, though the effect was stronger in continuation-base than in question-base stimuli. Hence our hypothesis that French continuation and question rises can be identified on the basis of rise shape, independent of L and H alignment, as well as target scaling and segmental duration, was confirmed by the results.

Despite the fact that our results provide further support to the importance of dynamic pitch information in the perception of intonation contrasts, they cannot be easily accounted for by the ToCG hypothesis, stating that contour shape might induce a different tonal alignment perception. According to ToCG, convex stimuli should translate into a later perceived H alignment. In French, though, H alignment is not significantly different between French question and continuation rises in production [28], hence alignment cannot be the basis of the perception difference. A more dynamic interpretation of the perception results is hence in order.

Our findings are, on the other hand, in line with those by Grundstrom [15], suggesting that concave interpolation is a feature of perceived question intonation in French. The author also noticed, though, that greater levels of identification were guaranteed by the rise being produced at a sufficiently high F0 level. Given that both duration and tonal target values were rendered ambiguous between the two base stimuli, our results point to the fact that the listener, when deprived of clear alignment and scaling information, can employ dynamic F0 information in order to distinguish between the two contours. Note also that stimuli at step 0, which were only durationally modified in order to be ambiguous between the two modalities, were still clearly identified as belonging to either one of the modalities, possibly thanks to keeping the original scaling values (i.e., a higher final H for questions). This suggests that duration itself might not be a crucial cue for identification of this contrast in French.

Morever, Base type had an effect on the manipulation, given that our statistical analyses showed an interaction between Step and Base. Specifically, the shift from continuation to question and vice-versa was not homogeneous, given the tendency for continuation-base stimuli to produce a later shift (9.63 step as cross-over point), with a softer slope in the response curve which is visible in Fig. 3. We can speculate that this might have been due to specific intensity and/or spectral characteristics of the base stimuli. As for intensity, Rossi [29] pointed out the need for this acoustic feature to reach specific values in order to induce the perception of specific intonation contrasts. Grundstrom [15] also showed that greater intensity values towards the end of the rise continuum increased question identification, to the exception of stimuli characterized by higher F0 ranges. Note that intensity and F0 do also interact in the perception of lexical tone in Mandarin [32]. Thus, given that neither intensity nor spectral cues were rendered neutral in our stimuli, they might have induced the observed interaction.

Finally, note that the slope of the response curves is rather shallow for both continua (though slightly less for the continuation-base continuum) hence not bearing the typical S-shape curve of categorically perceived stimuli. This might have been merely due to the use of very small F0 steps. Alternatively, this can be accounted for by the fact that rise shape might be a secondary cue relative to H target scaling and/or duration.

In sum, our results show that interpolation rise shape can act as a cue differentiating intonation contours in perception, in the absence of unambiguous tonal target alignment and scaling information. Hence, a complex integration of F0, intensity and possibly spectral dynamics might be at the basis of the perceptual representation of intonation, similar to lexical tone contrast. Future work will need to assess whether dynamic cues are employed only when primary cues are missing and/or in the presence of noise.

4. CONCLUSION

Our study confirms sensitivity on the part of French listeners to dynamic changes in intonation rises, and its use in understanding speaker's pragmatic intentions. Stimuli varying in degree of convexity within an F0 rise were either identified as continuations or as questions, in absence of unambiguous target scaling, tonal alignment and durational cues. Future work will need to determine whether dynamic information is further reanalysed in terms of perceived static targets, with specific alignment and scaling values, or if the representation of intonation is itself dynamic in nature.

5. REFERENCES

[1] André C., Ghio A., Cavé C., Teston B. 2003. PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception. *Proc. of 15th ICPhS*, Barcelona, 1421-1424

[2] Arvaniti A., Ladd, D., Mennen, I. 2006. Tonal Association and Tonal Alignment: Evidence from Greek Polar Questions and Contrastive Statements. *Language and Speech 49*. (Pt 4):421-50.

[3] Arvaniti, A., Ladd, D., Mennen, I. 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, 26: 3-25.,

[4] Barnes, J., Veilleux, N., Brugos, A., Shattuck-Hufnagel S. 2012. Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology*. 3(2), 335-383.

[5] Bates, D., Bolker, B., Walker, S. 2016. Package 'lme4', version 1.1- 12, linear mixed effects models using 'Eigen' and S4.

[6] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program], specified versions 2009.

[7] Crystal, D. 1969. Prosodic systems and Intonation in English, Cambridge: Cambridge University Press.

[8] D'Imperio, M. 2000. *The role of perception in defining tonal targets and their alignment*, Unpublished Ph. D. thesis, the Ohio State University.

[9] D'Imperio, M., House, D. (1997), "Perception of questions and statements in Neapolitan Italian.", *Proc. of Eurospeech'97*, Rhodes, 251-254.

[10] D'Imperio, M., Cangemi F. 2009. The interplay between tonal alignment and rise shape in the perception of two Neapolitan rising accents. Poster presented at at PaPI, Gran Canaria.

[11] D'Imperio, M., Gili Fivela, B., Niebuhr, O. 2010. Alignment perception of high intonational plateaux in Italian and German. *Proc. of Speech Prosody 5*, Chicago.

[12] Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Jun, S-A., Martin, P., Meisenburg, T., Rialland, A., Sichel-Bazin, R., Yoo, H. 2015. Chapter 3 – Intonational phonology of French: Developing a ToBI system for French. In Frota, S., Prieto P. *Intonation in Romance.* Oxford University Press.

[13] Delattre, P. 1966. Les Dix Intonations de base du français. *French Review*, 40(1), 1-14.

[14] German, J., D'Imperio, M. 2016. The Status of the Initial Rise as a Marker of Focus in French. *Language and Speech*. 59(2), 165–195.

[15] Grundstrom, A. 1973. L'intonation des questions en français standard. Studia Phonetica 8: Interrogation et intonation.19-49

[16] Gussenhoven C., Rietveld T. 1998. On the speaker-dependence of the perceived prominence of F0 peaks. *Journal of Phonetics*. 26, 371—380.

[17] 't Hart, J. 1991. F0 Stilization in speech: Straight lines versus parabolas. *The Journal of the Acoustical Society of America*. 90 (6), 3368-3370.

[18] 't Hart, J., Collier, R., Cohen, A. 1990. *A perceptual study of Intonation. An experimental-phonetic approach to speech melody.* Cambridge: Cambridge University Press.

[19] 't Hart, J.1981. Differential sensitivity to pitch distance, particularly in speech. *The Journal of the Acoustical Society of America*. 69 (3), 811-821.

[20] House, D. 1990. *Tonal Perception in Speech*. Lund: Lund University Press.

[21] Knight, R.A., Nolan, F. 2006. The effect of pitch span on intonational plateaux. *Journal of the International Phonetic Association.* 36(1), 21-38.

[22] Niebuhr, O. 2003. Perceptual study of timing variables in F0 peaks. *Proc. of 15th ICPhS*, Barcelona, 1225-1228

[23] Niebuhr, O., D'Imperio M., Gili Fivela B., Cangemi F. 2011. Are there "shapers" and "aligners"? Individual differences in signalling pitch accent category. *Proc. of 17th ICPhS,* Hong Kong, 120-123.

[24] O'Connor, J.D., Arnold, G.F. [1961]1973. *Intonation of colloquial English*. London: Longman.

[25] Petrone, C., D'Imperio, M. 2008. Tonal structure and constituency in Neapolitan Italian: evidence for the accentual phrase in statements and questions. *Proc. of Speech Prosody.* Campinas, 301-304.

[26] Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation*, Ph.D thesis, Massachusetts Institute of Technology.

[27] Post, B. 2002. *French Tonal Structures*. *Proc. of Speech Prosody*. 583-586.

[28] Regnier, M. (2014). *Prosodie des montées finales en français méridional*. Mémoire d'Orthophonie 2013-2014

[29] Rossi, M. 1979. Les configurations et interaction des pentes de F0 et de I. *Travaux de l'Institut de Phonetique d'Aix.* 6: 49-72.

[30] Trager, G. L., Smith, H. L. Jnr. 1951. *An outline of English Structure*. Washington: American Council of Learned Societies.

[31] Valtersson, E., Torreira, F. (2014). Rising intonation in spontaneous French: How well can continuation statements and polar questions be distinguished? In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proc. of Speech Prosody*. 785-789.

[32] Whalen, D., Xu, Y. 1992. Information for Mandarin Tones in the Amplitude Contour and in Brief Segments. *Phonetica*. 49(1):25-47.