Infants' perception of native and non-native pitch contrasts

Sónia Frota[1], Joseph Butler[1], Shuang Lu[2], Marina Vigário[1]

[1] Centre of Linguistics, University of Lisbon, Portugal
[2] School of Foreign Languages, Renmin University of China, China

sonia.frota@mail.telepac.pt, jbutler@fl.ul.pt, Shuanglu2013@outlook.com,
marina.vigario@mail.telepac.pt

Abstract

Infants' ability to distinguish between forms of phonetic variation in speech that are relevant to meaning is essential for their language development. Little is known about the developmental course of infants' perception of pitch contrasts, particularly in the presence of segmental variability which entails the ability to extract and generalize the contrastive patterns. Using single-bisyllabic utterances, in Experiment 1 we examined native discrimination of the statement (falling)/yes-no question (falling-rising) intonation contrast by European Portuguese (EP)-learning infants, and demonstrated that both 5-6 and 8-9 month-old infants were able to discriminate the contrast. Experiment 2 addressed the question whether the contrast between falling vs. falling-rising contours would also be perceived in segmentally varied non-native input. EP-learning infants' perception of the lexical distinction between Mandarin Chinese Tone 1+Tone 4 and Tone 1+Tone 2, with overall similar contour shapes but different implementation of the falling/rising patterns throughout the segmental string, was examined. Infants failed to discriminate the non-native pitch contrast, both at 5-6 months and 8-9 months, suggesting that language-specific perception for pitch, and for the tone/intonation distinction, emerges as early as 5 months of age.

Index Terms: Infant perception, development, pitch, intonation, statements, questions, lexical tone, native, non-native

1. Introduction

Infants' ability to distinguish between forms of phonetic variation in speech that are relevant to meaning is essential for their language development. Studies into the acquisition of sound categories have shown a developmental change before the infant's first birthday, whereby discrimination abilities become specifically attuned to the native language [1], [2]. This general developmental pattern may be modulated by perceptual assimilation differences and by the phonetic salience of both non-native and native contrasts [3], [4]. In this paper we are concerned with infants' perception of both native (intonation), and non-native (lexical tone) pitch contrasts and their developmental course in the first year of life.

The few studies on the development of infants' perception of prosodic categories have mostly investigated distinctions at the lexical level. These studies revealed differences in discrimination abilities for native stress, tone and lexical pitch accent contrasts: for stress a protracted development of discrimination is reported, whereas for lexical pitch accent and tone precocious discrimination abilities were found (see [5],

for a review). Importantly, segmental variability seems to affect early perception of some prosodic contrasts only (e.g., stress is negatively affected [6], but not pitch accent [7]).

Prior research looking at the perception of lexical tone distinctions has observed that learners of tone languages, such as Mandarin or Cantonese, show stable discrimination of acoustically distinct lexical tones throughout the first year, as early as 4 months. By contrast, learners of non-tone languages, such as English, French or Dutch, show a decline in tone discrimination from 4 to 9 months of age, even in the presence of a very salient tonal contrast [8], [9], [10], [11]. In these studies, tone perception was examined in the absence of segmental variability. Interestingly, infants' ability to perceive native tone contrasts in the presence of segmental variation seems to develop later, by 8-11 months [12].

Little is known about the developmental course of infants' perception of intonation. A few studies have started to examine the discrimination of native sentence type intonation contrasts by infants, mostly focusing on the English language [13], [14], [15], [16]. From the mixed pattern of findings reported, it is unclear how the perception of sentence type intonation unfolds during the first year of life. A recent study, examined infants' perception of statement (falling)/yes-no question (falling-rising) intonation in European Portuguese (EP), a language that, unlike English, marks this sentence type contrast only by prosodic means [5]. This study demonstrated successful infant discrimination of the intonation contrast throughout the first year, suggesting precocious discrimination abilities for intonation similar to those reported for lexical pitch, in the presence of segmental variation. However, this precocious ability of EP-learning infants was shown not to extend to other types of native intonation contrasts, suggesting differing developmental trajectories for different kinds of pitch contrasts [17].

Building on the findings in [5], non-native discrimination of EP falling versus falling-rising intonation was investigated in English-learning and Basque-learning infants. The former failed, whereas the latter succeeded, showing that infants' perception of intonation contrasts is already language-specific at 4-5 months of age [18].

To explore infants' perception of native and non-native pitch contrasts, we revisited EP-learning infants' perception of statement/yes-no question intonation (Experiment 1), and asked whether the salient contrast in pitch (falling/low versus falling-rising/high) would also be perceived in segmentally varied non-native input. This was tested (Experiment 2) using the Mandarin Chinese contrast between Tone 1+Tone 4 (akin to the final fall of statements) and Tone 1+Tone 2 (akin to the final fall-rise of questions) [19]. Given the overall similar contour shapes, similar patterns of discrimination might be

expected across experiments. However, if infants' perception is indeed influenced by aspects of language experience, such as different distributions of the falling and rising contours across the utterance, different discrimination patterns should emerge.

2. Experiment 1

Experiment 1 investigated EP-learning infants' ability to discriminate the native falling versus falling-rising intonation contrast that characterizes the difference between statements and yes-no questions. The use of stimuli with segmental variability required the ability to extract and generalize the contrastive pitch patterns.

2.1. Participants

Forty typically developing infants from monolingual EP homes participated in this study, split into two age groups; twenty younger (8 female, mean age 5 months 29 days, range 5 months 3 days – 6 months 23 days) and twenty older (10 females, mean age 8 months 12 days, range 7 months 11days – 9 months 29 days). Eight additional infants (three younger, five older) were rejected due to fussiness (5), a parent who was not European Portuguese (1), an older, autistic sibling (1), and software error (1).

2.2. Stimuli

Sixteen pseudo-words (malo, lemo, mela, rono, rano, nurra, nirra, lamo, milo, mola, norro, reno, nerra, rina) were used to create statement and question segmentally varied single-prosodic word utterances. The pseudo-words were bisyllabic, all sonorant sequences with initial stress, and were produced by a female native speaker in Infant Directed Speech (IDS). Consonants with varying frequency in the language were used, so that the stimuli included less common appearing syllables (e.g., 're' or 'ro'). Different pseudo-words were used for the habituation and test phase.
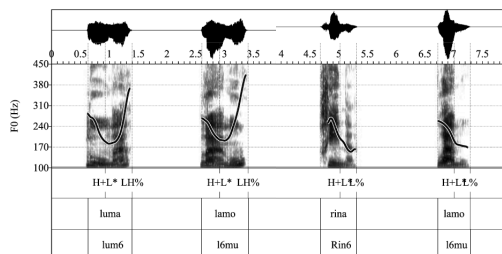

Figure 1: *Illustrative contours of question (left) and statement (right) stimuli, with prosodic labeling, orthographic and phonetic transcription (SAMPA).*

EP yes-no questions show an accentual fall and a rising boundary (H+L* LH%), whereas statements show a accentual fall and low boundary tone (H+L* L%), as in Figure 1. Acoustic analysis carried out on the stimuli revealed that the two types of stimuli are indeed distinguished by the boundary tone difference (L% for declaratives, final F0 range = -24.75 Hz, final F0 value = 163 Hz; for interrogatives, final F0 range = 191.56 Hz, final F0 value = 380 Hz), together with longer duration for yes-no questions. Detailed acoustic analysis of the stimuli is found in [5].

2.3. Procedure

A modified version of the visual habituation paradigm [20], [7] was implemented, using LOOK software [21]. Infants were seated on their caregivers lap, facing a computer monitor displaying an attractive image. Once the infant fixated the image, auditory stimuli were presented at the same time a red and black checkerboard display was shown. The experiment had two phases: habituation and test. In the habituation phase, infants were presented with either statements or questions (counterbalanced). The habituation trials were sound files with eight pseudo words, and the duration of 16 seconds. After the end of the sound file, the monitor reverted to a colorful image, and the next trial began once the child fixated the image. When average listening time to the last four habituation trials heard was <60% of the average listening time to the first four habituation trials, the test phase began. In the test phase, infants were presented with two test trials, one "familiar" (as the habituation) and one "novel" (different to the habituation), with order of test trials counterbalanced. Infants' listening times to familiar and novel trials were statistically compared. If infants demonstrated a sensitivity to the prosodic contrast, the listening times to the novel trials should be longer.

2.4. Results and Discussion

Average listening times to familiar and novel test trials by age group are presented in Figure 2 (left). Both groups display longer looking times to the switch test trials. Overall, 18 out of 20 younger and 17 out of 20 older infants had longer listening times to the switch test trials.
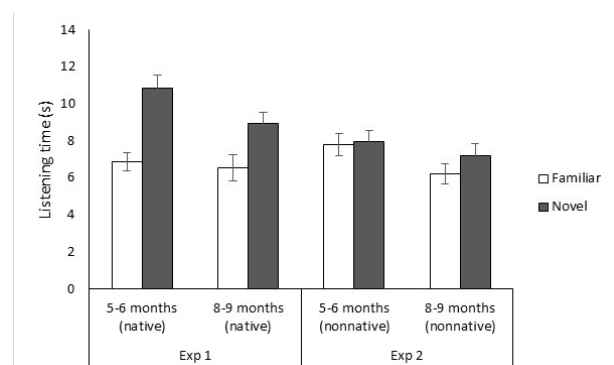

Figure 2: *Average listening times to the familiar/novel test trials, across the two age groups, for Experiment 1 (left) and Experiment 2 (right). Error bars indicate the standard error of the mean.*

A repeated measures ANOVA with a within-participant factor of trial type (familiar, novel) and two between-participant factors of age group (younger, older) and habituation condition (statement, question) revealed a significant main effect of trial type only, $[F(1,36) = 54.18, p < .001, \eta^2 = .6]$. No other significant effects were found (age group $[F(1,36) = 2.13, p = .15, \eta^2 = .06]$; habituation condition $[F(1,36) = 2.02, p = .16, \eta^2 = .05]$, or interactions – trial type x age group $[F(1,36) = 3.29, p = .08, \eta^2 = .08]$; all other interactions $[F(1,36) < 1]$). Paired t-tests for the two age groups separately revealed significant differences between trial type for both younger $[t(19) = 6.1, p < .001, d = 1.474]$, and older $[t(19) = 4.42, p < .001, d = 0.816]$ age groups.

Thus, EP-learning infants are able to successfully discriminate statements and yes-no questions, in the presence of segmental variability. These findings support the view that pitch contrasts, at least those based on pitch height and/or direction, are especially salient to infants [22], [17], in particular if the native language uses them in consistent ways as (the sole or main) cues to linguistic categories [5], [18].

3. Experiment 2

Experiment 2 investigated whether EP-learning infants are able to discriminate a falling versus falling-rising pitch contrast in a non-native input. Infants were presented with the Mandarin Chinese Tone 1+Tone 4 versus Tone 1+Tone 2 contrast, which presents a similar contour shape to the statement/yes-no question intonation contrast in EP albeit with a differing implementation of the falling and falling-rising patterns throughout the segmental string.

3.1. Participants

Forty typically developing infants from monolingual EP homes participated in this study, split into two age groups; twenty younger (8 female, mean age 5 months 25 days, range 5 months 2 days – 6 months 19 days) and twenty older (10 females, mean age 8 months 21 days, range 7 months 13 days – 10 months 8 days). In addition, nine infants (five younger, four older) were rejected due to fussiness (5), bilingual (2), an older, autistic sibling (1), and software error (1).

3.2. Stimuli

Sixteen pseudo-words or actual words in Mandarin Chinese (kelu, fali, badi, gudi, boli, godu, bali, kefu, badu, pabi, balu, goli, disu, godi, keli, guli) were used to create bisyllabic Tone 1+Tone 2 and Tone 1+Tone 4 segmentally varied single-prosodic word utterances. The degree of segmental variability was similar between the Mandarin Chinese (13 phonemes) and the EP stimuli (11 phonemes). Only consonants and vowels that also exist in EP were included in the Mandarin stimuli. Similarly to the EP pseudo-words, the Mandarin stimuli included consonants with varying frequency in EP. Thus the segmental make up of the non-native stimuli was controlled to show similar features to the native stimuli used in Experiment 1. The utterances were produced by a female native speaker of Mandarin Chinese in IDS. Different (pseudo)words were used for the habituation and test phase. The sound files of the stimuli are available at http://labfon.letras.ulisboa.pt/babylab/Infants_Perception/non_native.htm.

As noted for other languages [19], Tone 4 approximates falling (declarative) contour and Tone 2 rising (question) contour in EP intonation. Thus, the Mandarin and the EP pitch contours show overall similar contour shapes. Moreover, the direction of the boundary pitch movement, a crucial cue for the sentence type distinction in EP, is the same (Figure 3).

Similarly to EP, the Mandarin Chinese stimuli differed in relation to final pitch height and direction (Tone 1+Tone 4, final F0 range = -103 Hz, final F0 value = 205 Hz; Tone 1+Tone 2, final F0 range = 35 Hz, final F0 value = 284 Hz, p < .001) and longer duration for rising (p < .01), but not in first F0 peak height (306 Hz) or first syllable F0 range (11 and 10 Hz). Notably, the complexity of the native and non-native contours is similar, both phonetically (one turning point in the

falling, two in the falling-rising) and phonologically (including four or three tones each).
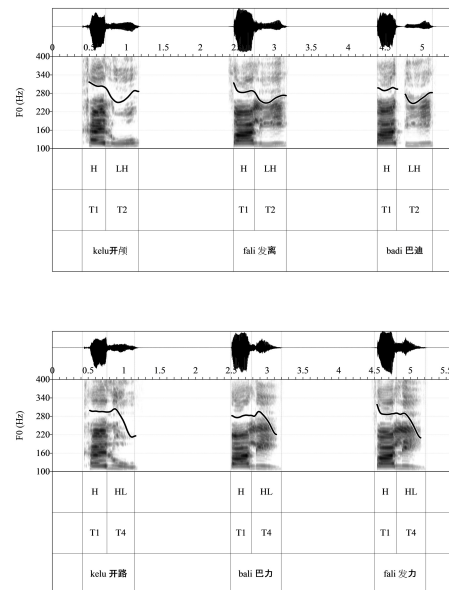

Figure 3: *Illustrative contours of Tone 1+Tone 2 (top) and Tone 1+Tone 4 (bottom) stimuli, with prosodic labeling, and orthographic transcription.*

These similarities, however, co-exist with prosodic differences. In particular, there are different distributions of the falling and rising contours: in EP, the first syllable shows falling pitch in both statements and questions, whereas in Mandarin the first syllable always shows high pitch (Tone 1); in EP, the pitch contrast in the second syllable is between low (L%) and rising pitch (LH%), whereas in Mandarin a falling (HL, due to Tone 4) and falling-rising (HLH, due to the transition from Tone 1 to Tone 2) difference obtains. These differences arise from the different nature of tune versus tone: the tune spreads over the bisyllabic utterance; the tone has syllable as its domain. In addition, the pitch range difference in the second syllable is reversed: in EP, the rising boundary (question) has wider pitch range than the falling (statement); in Mandarin, the falling tone (Tone 4) shows wider pitch range than the rising tone (Tone 2).

3.3. Procedure

The procedure for Experiment 2 was identical to that used in Experiment 1, expect the infants were habituated and tested with Mandarin Chinese Tone1+Tone4 or Mandarin Chinese Tone 1+Tone 2, rather the statement/yes-no question EP intonation contrast.

3.4. Results and Discussion

Average listening times to familiar and novel test trials by age group are shown in Figure 2 (right). Both groups display similar listening times to the familiar and novel test trials (younger familiar mean – 7.81, SD – 2.65, novel mean – 7.97, SD – 2.59; older familiar mean – 6.21, SD 2.44, novel mean –

7.21, SD – 2.94). Overall, 10 out of 20 younger and 11 out of 20 older infants listened longer to the novel test trials.

A repeated measures ANOVA with a within-participant factor of trial type (familiar, novel) and two between-participant factors of age group (younger, older) and habituation condition (Tone 1+Tone 2, Tone 1+Tone 4) revealed no significant main effect of trial type [F(1,36) < 1], age group [F(1,36) = 3.02, p = .09, η^2 = .08], or habituation condition [F(1,36) < 1]. There were no significant interactions (all interactions, F(1,36) < 1). Paired t-tests for the two age groups separately revealed no difference between familiar and novel test trials for both younger [t(19) = .27, p = .79, d = 0.06], and older [t(19) = .86, p = .390, d = 0.241] age groups.

Unlike in Experiment 1, EP-learning infants were not able to successfully discriminate the Mandarin Chinese lexical tone contrast, in the presence of segmental variability, both at 5-6 and 8-9 months of age.

To further inspect the different pattern of findings between Experiment 1 and Experiment 2, the results from both experiments were compared. Given the same sample size (n = 20 per group), a repeated measures ANOVA was used with a within-participant factor of trial type (familiar, novel) and two between-participant factors of language (EP, Mandarin) and age group (younger, older). Results revealed a significant main effect of trial type [F(1,76) = 18.71, p < .001, η^2 = .2] and age group [F(1,76) = 5.09, p < .05, η^2 = .06], and an almost significant effect of language [F(1,76) = 3.77, p = .056, η^2 = .05]. Crucially, there was a significant interaction between trial type and language [F(1,77) = 26.79, p < .001, η^2 = .26], reflecting the fact that EP-learning infants responded differently to native and non-native novel trials, but not to familiar trials (see Figure 1). All other effects and interactions were not significant (trial type by age group [F(1,76) = 3.4, p = .07, η^2 = .04], language by age group [F(1,76) < 1], status by language by age group [F(1,76) < 1]).

These findings do not support the view that pitch contrasts, namely those based on pitch height and/or direction, are especially salient to infants who thus show precocious discrimination abilities for such contrasts [5], [22], [17]. The present findings, however, add to recent research highlighting the role of language experience on infants' perception of pitch very early on in development.

4. General Discussion

In this paper, we examined 5- to 6- and 8- to 9-month-old EP learning infants' perception of a native (Experiment 1) and non-native (Experiment 2) pitch contrast, in the presence of segmental variability.

The native contrast was the distinction between statement and yes-no question sentence types, thus an intonation distinction. The crucial cue for this distinction has been reported to be a boundary tone difference reflected in pitch direction and pitch height (respectively, falling/low versus rising/high) [5], [23], [24]. In Experiment 1, EP-learning infants successfully discriminated the native intonation contrast at both ages. The non-native contrast was the difference between Mandarin Chinese Tone 1+Tone 4 and Tone 1+Tone 2, thus a lexical tone distinction. Importantly, the lexical tone distinction is realized with a similar contour shape to the intonation contrast, with Tone 4 resembling falling statement intonation and Tone 2 rising question intonation [19]. However, despite the similar contour shapes,

that could arguably be approximated to similar (intonation) functions by non-tone language learners, in Experiment 2 EP-learning infants failed to discriminate the Mandarin stimuli at both ages. Given that the natural segmental content of the native and non-native stimuli was controlled to show similar features (namely Mandarin segments that also exist in EP and word-initial and word-medial consonants with varying frequency of occurrence in the language, thus including less commonly appearing syllables), this failure to categorize the two different sets of contours suggests that EP infants could distinguish between native and non-native falling/low versus rising/high pitch, as early as at 5 months of age. In fact, the native and non-native contours, although similar, show different distributions of the falling and rising pitch across the bisyllabic utterances. These distributions seem to arise from the difference in nature between the tune, which takes the bisyllabic utterance as its domain, and the tone, which has the syllable as its domain. Furthermore, there are pitch range differences that characterize the contrast across the two languages, with a wider pitch range of the falling contour in Mandarin and a wider pitch range of the rising contour in EP.

Therefore, the present findings support the conclusion that infants' perception of pitch contrasts is influenced by language experience, and that language-specific perception for pitch, and for the tone/intonation distinction, emerges as early as 5 months of age, and earlier than for vowels, consonants or stress [6], [10]. Further research is needed, however, to better define the aspects of infants' language particular experience behind this precocious sensitivity to pitch differences across languages. If the interpretation put forward here is on the right track, we might expect EP-learning infants to have less difficulties in perceiving the contrast between falling and rising contours in a lexical pitch accent language, like Japanese, which has bisyllabic words distinguished by HL and LH contours [7]. Unlike tone, but similarly to intonation, the lexical pitch accent contrast has the single bisyllabic-word utterance as its domain. The results from a follow-up study we are currently running confirm this prediction, as EP-learning infants are indeed better at discriminating the Japanese pitch contrast than the Mandarin tone contrast.

5. Acknowledgements

6. References

[1] P. K. Kuhl, "Early language acquisition: Cracking the speech code", *Nature Reviews Neuroscience*, vol. 5, pp. 831-843, 2004.

[2] J. R. Saffran, J. F. Werker and L. A. Werner, "The infants' auditory world: Hearing, speech, and the beginnings of language", in D. Kuhn and M. Siegler (eds.), *The 6th edition of the handbook of child psychology*. New York: Wiley, pp. 58-108, 2006.

[3] C. C. Best and G. W. McRoberts, "Infant perception of non-native consonant contrasts that adults assimilate in different ways", *Language and Speech*, vol. 46, 183-216, 2003.

[4] C. R. Narayan, J. F. Werker and P. S. Beddor, "The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place

discrimination", *Developmental Science*, vol. 13, no. 3, pp. 407-420, 2010.

[5] S. Frota, J. Butler and M. Vigário, "Infants' perception of intonation: Is it a statement or a question?", *Infancy*, vol. 19, no. 2, pp. 194-213, 2014.

[6] K. Skoruppa, F. Pons, L. Bosch, A. Christophe, D. Cabrol and S. Peperkamp, "The development of word stress processing in French and Spanish infants", *Language Learning and Development*, vol. 9, no. 1, pp. 88-104, 2013.

[7] Y. Sato, Y. Sogabe and R. Mazuka, "Development of hemispheric specialization for lexical pitch-accent in Japanese infants", *Journal of Cognitive Neuroscience*, vol. 22, no. 11, pp. 2503-2513, 2009.

[8] K. Mattock and D. Burnham, "Chinese and English infants' tone perception: Evidence for perceptual reorganization", *Infancy*, vol. 10, no. 3, pp. 241-265, 2006.

[9] K. Mattock, M. Molnar, L. Polka and D. Burnham, "The development course of lexical tone perception in the first year of life", *Cognition*, vol. 106, no. 3, pp. 1367-1381, 2008.

[10] H. H. Yeung, K. H. Chen and J. F. Werker, "When does native language input affect phonetic perception? The precocious case of lexical tone", *Journal of Memory and Language*, vol. 68, no. 2, pp. 123-139, 2013.

[11] L. Lui and R. Kager, "Perception of tones by infants learning a non-tone language", *Cognition*, vol. 133, no. 2, pp. 385-394, 2014.

[12] R. Shi, "Contextual variability and infants' perception of tonal categories", *Chinese Journal of Phonetics*, vol. 2, pp. 1-9, 2010.

[13] C. T. Best, A. Levitt and G. W. McRoberts, "Examination of language-specific influences in infants' discrimination of prosodic categories", in *Proceedings of the XIIth International Congress of Phonetic Sciences,* vol. 4. Aix-en-Provence: Université de Provence, France, pp. 162-165, 1991.

[14] M. Soderstrom, E. Ko and U. Nevzorova, "It's a question? Infants attend differently to yes/no questions and declaratives", *Infant Behavior and Development*, vol. 34, pp. 107-110, 2011.

[15] S. Geffen and T. Mintz, "Seven-month-olds discrimination of statements and questions", in A. K. Biller, E. Y. Chung and A. E. Kimball (eds.), *Proceedings of the 36th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 2011.

[16] S. Geffen, "When and how infants discriminate between declaratives and interrogatives. Ph.D. Dissertation, University of Southern California, 2014.

[17] J. Butler, M. Vigário and S. Frota, "Infants' perception of the intonation of broad and narrow focus", *Language Learning and Development,* pp. 1-13, 2015.

[18] M. Sundara, M. Molnar and S. Frota, "The perception of boundary tones in infancy", in The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, UK: the University of Glasgow, 2015.

[19] B. Braun and E. K. Johnson, "Question of tone 2? How language experience and linguistic function guide pitch processing", *Journal of Phonetics*, vol. 39, no. 4, pp. 585-594, 2011.

[20] C. L. Stager and J. F. Werker, "Infants listen for more phonetic detail in speech perception than in word learning tasks", *Nature*, 388 (6640), pp. 381-382, 1997.

[21] K. Meints and A. Woodford, *Lincoln Infant Lab Package 1.0: A new programme package for IPL, Preferential Listening, Habituation and Eyetracking*, 2008. [WWW document: Computer software & manual]. URL: http://www.lincoln.ac.uk/psychology/babylab.htm

[22] R. A. H. Bion, S. Benavides-Varela and M. Nespor, "Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences", *Language and Speech*, vol. 54, no. 1, pp. 123-140, 2011.

[23] S. Frota, "Nuclear Falls and Rises in European Portuguese: A Phonological Analysis of Declarative and Question Intonation", *Probus*, vol. 14 (Special issue on intonation in Romance, ed. by J. Ignacio Hualde), pp. 113-146, 2002.

[24] I. Falé and I. H. Faria, "Intonational contrasts in EP: a Categorical Perception approach", in *Proceedings of the 9th European Conference on Speech Communication and Technology*. Lisbon, pp. 1705-1708, 2005.