# Pitch range variation in child affective speech

I. Grichkovtsova & I. Mennen

Speech and Hearing Sciences QMUC, Edinburgh, UK igrichkowtsova@qmuc.ac.uk

## Abstract

This study investigates pitch range variation in the affective speech of bilingual and monolingual children. Cross-linguistic differences in affective speech may lead bilingual children to express emotions differently in their two different languages. A cross-linguistically comparable corpus of 6 bilingual Scottish-French children and 12 monolingual peers was recorded according to the developed methodology. The results show that the majority of children use pitch range measurements (overall level and span) to realize differences between some emotions. Monolingual children use analyzed acoustic parameters in a much more homogeneous way than bilinguals. Some results of bilingual children do not strictly correspond to those of monolinguals, and show bidirectional interference.

## 1. Introduction

The main objective of this study is to investigate the production of pitch range in the affective speech of bilingual and monolingual children cross-linguistically. Recent studies [1, 2] suggest that there is a number of available means to express emotions in speech (pitch range, rhythm, voice quality, etc), but their usage, level of importance and meaning vary in different languages. While there is some evidence on the cross-linguistic differences in the adult productions of affective prosody [3], their realization in child speech remains to be addressed. Another particularly interesting question concerns bilingual mastering of affective prosody. Bilinguals communicate their emotions in two different languages, and they have to use acoustic means related to affective speech appropriately. Given that bilinguals may have a wider range of means and ways to express emotions, their affective realization may not strictly correspond to those of monolinguals in each of their two languages. This phenomenon has been attested in bilingual studies of other phonetic aspects [4, 5]. Understanding of child monolingual and bilingual usage of affective prosody will be specifically important for theories of phonetic learning, and may have implications for second language acquisition theories in general.

### 2. Methodology

Though spontaneous speech is claimed to contain the most authentic emotions, its collection is ethically problematic, especially with children. Thus, this study records acted speech which does not suffer the same ethical problems. Moreover, acting of emotions can be presented as a game to children, so that they can be more naturally involved in the expression of emotions. Visual materials were developed in a realistic manner together with a professional illustrator and based on the research of facial expressions. They represent a child expressing four emotions (happiness, sadness, anger and fear); the child's gender and age are adapted to that of the recorded children. The randomized cards have 16 repetitions for each emotion. One token utterance was selected with a similar sound and prosodic structure for English and French: "I see a banana there./Je vois une banane, moi." The subject is instructed to say the token utterance in the same way as the child at the picture, thus the child is playing emotions through association with the drawing.

Neutral was not represented as a facial expression card. Neutral utterances were recorded in a separate test at the very beginning of the recording session. In this test, the child was instructed to say the token utterance (I see \* there./Je vois \*, moi.) with a word from a pile of cards (picture naming); 16 repetitions of the original word (banana/banane) were mixed among the others. Later all the utterances with the original token wording were selected for analysis. A number of utterances was rejected due to noise, hesitation or incorrect wording (10%). Bilingual children were recorded in two sessions (one for each language) by different experimenters in order to control for language mode and code-switching; there was a period of about two weeks between the sessions. This methodology allowed us to record a cross-linguistically comparable corpus of 6 bilingual Scottish-French children and 12 monolingual peers, average age - 8 years old.

### 2.1. Participants

This paper presents data for four bilingual and eight monolingual children, taken from the recorded corpus. Bilingual children (children 1-4) were born in Edinburgh, Scotland, with a French mother and a Scottish father, except child 3 who has both French parents. They go to Edinburgh mainstream schools, and attend the same French afternoon school twice a week. There is some variation in the amount of French input the children are receiving. English is the stronger language for all the children, as it is the language of their school, friends, and the country. French monolingual children (children 5-8) were recorded in Normandy, France. Scottish monolingual children (children 9-12) were recorded in Edinburgh, with both Scottish parents.

#### 2.2. Recording

All children were recorded with a TASCAM DA-P1 portable DAT recorder either in the recording studio or in a quiet school room. The speech was digitized and stored at 44100 Hz. The analysis was performed on the speech downsampled to 22050 Hz.

#### 2.3. Pitch range measurements

Following Ladd [6] we describe pitch range using two partially independent measures of variation: overall level and span. Overall level refers to the "height" of a speaker's voice, span refers to the width of pitch frequencies covered by a speaker (how big the excursions are). These measures have been shown [7] to give better and perceptually relevant results, than the more commonly used measures, such as statistical moments (mean, standard deviation, difference between maximum and minimum  $F_0$ , etc.).

Using *wavesurfer* software, acoustic measurements of pitch range variation were taken. All the utterances were manually labeled at the following pitch points:  $IF_0$  (initial value of  $F_0$ ), H1 (the peak of the first accented word), V (valley, the lowest point between two peaks), H2 (the peak of the main accented word main accent), FL (final low - the lowest value after the peak). In cases where there was no well defined peak, an absolute  $F_0$ max value was taken near the accented syllable.  $F_0$  values were then extracted automatically to data files with the help of ESPS algorithm, and again checked manually for any pitch perturbation or voice quality errors. In our measurements, FL point was taken as the overall level, and the difference of H2 and FL, as the span. All measures were first expressed in semitones.

As pitch range usage is well known for high individual differences, it was decided to normalize the pitch values, based on the models suggested in Ref. [8]. For each measurement,  $F_0$ values were transformed to Z scores according to the following formula:

$$Z = \frac{X - \mu_{\text{neutral}}}{\sigma_{\text{neutral}}},$$

where X is a measured  $F_0$  value for a given affective state and a given speaker in semitones, and  $\mu_{neutral}$  and  $\sigma_{neutral}$  are the mean and the standard deviation of the corresponding  $F_0$  values for neutral utterances of the same speaker. Z-score gives the distance between the measured value X and the mean  $\mu_{neutral}$  as a number Z times the standard deviation  $\sigma_{neutral}$ . Note that in the case of bilingual children the values of  $\mu_{neutral}$  and  $\sigma_{neutral}$ for one language are different of those for the other language.

## 3. Results

Results on pitch range - overall level and span - for four bilingual and eight monolingual children, comparing their emotions and languages are depicted in Figs. 1, 2 and 3. Separate multivariate ANOVAs were carried out for each child. Two factors were considered: EMOTION (five levels: anger, fear, sadness, happiness, and neutral) and LANGUAGE (two levels: English and French) for bilinguals; and one factor - EMOTION (the same five levels) for monolinguals. The dependent variables were pitch span and pitch level. Their statistical significance is shown in Table 1. There is no significant effect of factor LANGUAGE for the dependent variable level of child 1 and for span of child 2, and of factor EMOTION for the dependent variable span of child 6. In all the other cases, children use pitch range parameters to realize emotions. Bilingual children also use these parameters to differentiate some emotions in their two languages. Children 2, 3 and 4 separate happiness in their two languages in span, children 1 and 2 separate fear in span, and child 1 separate anger in span. Children 3 and 4 separate sadness and happiness in level. All the other emotions overlap in the two languages in their realizations. Thus, happiness is the most differentiated emotion by bilingual children. Post hoc tests were run to study the differences between particular emotions.

Table 1: MANOVA results for dependent variables - pitch span and level. The value is significant at p < 0.05. The first column includes the child number and factor (L - language, E - emotion, I - interaction).

I - Incraction).		
Child	Span	Level
1L	F(1,146)=42 p < 0.000	$F(1,146)=0.002 \ p < 0.965$
1E	F(4,146)=19 p < 0.000	F(4,146)=178 p < 0.000
1I	F(4,146)=13 p < 0.000	F(4,146)=1.2 p < 0.305
2L	F(1,130)=0.0 p < 1.0	F(1,130)=27.8 p < 0.000
2E	F(4,130)=16 p < 0.000	F(4,130)=66 p < 0.000
2I	F(4,130)=11 p < 0.000	F(4,130)=6 p < 0.000
3L	F(1,148)=27 p < 0.000	F(1,148)=67 p < 0.000
3E	F(4,148)=35 p < 0.000	F(4,148)=39 p < 0.000
3I	F(4,148)=11 p < 0.000	F (4,148)=18 $p < 0.000$
4L	F(1,154)=9 p < 0.004	F(1,154)=303 p < 0.000
4E	F(4,154)=19 p < 0.000	$F(4,154)=107 \ p < 0.000$
4I	F(4,154)=7 p < 0.000	$F(4,154)=55 \ p < 0.000$
5E	F(4,76)=7 p < 0.000	F(4,76)=62 p < 0.000
6E	F(4,64)=0.8 p < 0.562	F(4,64)=2.9 p < 0.028
7E	F(4,76)=15 p < 0.000	$F(4,76)=35 \ p < 0.000$
8E	F(4,76)=22 p < 0.000	F(4,76)=18 p < 0.000
9E	F(4,63)=3.4 p < 0.014	$F(4,63)=27.1 \ p < 0.000$
10E	F(4,71)=29.2 p < 0.000	F(4,71)=24.6 p < 0.000
11E	F(4,75)=11.2 p < 0.000	$F(4,75)=99.2 \ p < 0.000$
12E	F(4,78)=92.7 p < 0.000	$F(4,78)=49.5 \ p < 0.000$

Pitch level - bilinguals



Pitch span - bilinguals





#### 3.1. Pitch range results

#### 3.1.1. Post hoc tests for span

Child 1 realizes happiness with the highest span than the other emotions in English, it is the only emotion which is significantly different. In French, anger and fear have the widest span, they are not significantly different from each other, but are from the other emotions. Sadness and neutral have the narrowest span. Child 2 differs from child 1, as she realizes fear with the widest span in English, which is statistically significant. It is followed by anger, sadness, happiness (not significantly different), and by neutral with the narrowest span. In French span, anger is significantly different from the other emotions with the widest span. Child 3 results show that happiness has the widest span in English, as for child 1. Happiness is followed by anger which is also significantly different from the other emotions. Neutral, sadness and fear have the narrowest span, and they do not differ. In French, anger has the widest span; the other emotions have a narrower span, and they do not differ much from each other. Child 4 has happiness with the widest span in English, it is significantly different from the other emotions. In French, anger has the widest span, it is followed by fear and happiness, and finally by sadness and neutral with the narrowest span.

French monolingual child 5 also has anger with the widest span, it is significantly different from the rest of emotions, except neutral which overlaps both with anger and the other emotions with the narrower span; fear has the narrowest span. Child 6 realizes anger and happiness with a higher span than the other emotions emotions, sadness has the narrowest span, but the differences are not significant. Child 7 separates only neutral from the other emotions with the significantly narrower span; happiness and anger have slightly higher span, but it is not significant. Child 8 follows the same realization pattern as child 7, only neutral is significantly different with the narrowest span.

Scottish monolingual child 9 has fear with the widest span, but it is significantly different only from sadness with the narrowest span, which also overlaps with the rest of emotions. Child 10 also has fear with the widest span, and it is significantly different from the rest of emotions; neutral has the narrowest span, and it overlaps only with sadness. Child 11 has neutral with the narrowest span which is significantly different from the other emotions; happiness has the widest span, but it is significantly different only from sadness and neutral. Child 12 realizes span in the following order: happiness (the widest span), fear, anger, sadness, neutral. Only anger and fear overlap with each other, in other cases all the emotions are significantly different.

#### 3.1.2. Post hoc tests for level

Post hoc results for child 1 in English level show that fear has the highest level. It is followed by sadness which is also significantly different from the other emotions. Finally, happiness, anger and neutral have the lowest similar level. French has exactly the same pattern. In English, child 2 realizes fear and sadness with the highest level, anger and neutral have the lowest similar level, and happiness is in between these two groups and significantly different. In French, fear, sadness and happiness have the highest level, they are followed by anger which overlaps with sadness. Neutral has the lowest level, and it is considerably different from the other emotions. In English, child 3 realizes only neutral significantly different from the other emotions with the lowest level. In French, happiness has the highest level, it is followed by anger, fear, sadness (not different from each other), and finally by neutral with the lowest level, and significantly different. Child 4 has fear with the highest level which is significantly different from the rest of emotions, except sadness. All the other emotions overlap in English. In French, fear has the highest level, it is followed by sadness and happiness (not significantly different), and by anger (not significantly different from happiness); neutral is the lowest.

Child 5 results show that fear has the highest level and is significantly different from the rest of emotions, it is followed by sadness and happiness, then by anger and neutral. Child 6 almost does not separate emotions by level, she has two groups - fear, neutral and sadness with higher level, and happiness and anger with lower level, but these differences are almost not significant. Child 7 differentiates fear from the rest of emotions with the highest level, the other emotions overlap. Child 8 has fear with the highest level, it is followed by sadness, happiness and neutral (sadness is not significantly different from fear) and finally, anger has the lowest level, it is significantly different from the other emotions.

Child 9 has fear with the highest level, it overlaps only with sadness. Happiness, anger and neutral have the lowest level and are not significantly different from each other. Child 10 separates fear from the other emotions with the highest level, it is followed by sadness and neutral, and finally by happiness and anger with the lowest level. Child 11 realizes fear with the highest level; it overlaps only with sadness; the other emotions have the lowest level and are not significantly different from each other. Child 12 separates fear from the other emotions with the highest level; fear is followed by sadness, which is also significantly different from the other emotions. The other emotions overlap.

## 4. Discussion

Statistical tests show that there is a considerable amount of variability in the child affective realizations. Nevertheless, it is possible to make some generalizations of the results. Though the pitch level tends to be higher in French than in Scottish English, the general realization pattern of level seems to be similar in the two languages: fear has the highest level, then go sadness, happiness and finally anger. Pitch span tends to be wider in Scottish English, and its realization pattern presents some differences: French monolingual children realized anger and happiness with the widest span, while Scottish children realize fear and happiness with the widest span. Sadness is realized with the narrowest span in the two languages. One particular difference between monolingual children is that Scottish children differentiate more significantly their emotions by pitch range, both level and span, than French children.

#### 4.1. Pitch span by bilinguals

Bilingual children in French realize emotions in a comparable way with monolingual French children. The main difference is that bilinguals show significant differences between emotions more consistently. Bilinguals have anger with the widest span in French, like monolinguals, and it is significant for all of them. Child 2, 3 and 4 follow the realization pattern found among monolinguals. Although child 1 realizes anger with the widest span as other bilinguals and monolinguals, she stands out among the other children by realizing fear with a wide span, close to anger, while all the other children realize fear with a narrow span; thus three bilingual children does not separate fear and sadness in their French by span. Happiness is realized with a wide span only by one bilingual child (4).

Bilingual children in English are consistent in significantly separating emotions by span, like Scottish monolingual children. Three children realize happiness with the widest span, and child 2 realizes fear - with the widest span, which is consistent with Scottish monolinguals. For all the children, the separation of the emotion with the widest span is significant and there is no overlap with any other emotion. This separation is stronger in bilinguals, than in monolinguals (only two out of four monolinguals do not make overlap of the emotion with the widest span). Neutral and sadness have a narrow span. Only one bilingual child (2) has fear with a wide span, which is also significantly different from sadness. The fear realization by three bilingual children show difference from Scottish monolingual children, who realize fear with a wider span, and who tend to separate fear from sadness.

Though there is a tendency to separate fear from sadness by Scottish children, it is still acceptable not to do it (as one monolingual child shows it). As French monolingual children never separate these two emotions, it looks like the majority of bilingual children choose to follow this realization in their English as well, especially as it not impossible in Scottish English. Another interesting peculiarity of bilingual children is that they are consistent in realizing significant differences between emotions (in a similar and even stronger way as monolingual Scottish children), and they preserve this feature in French. Thus, they show realization patterns found in French monolingual children, but with a stronger significance.

#### 4.2. Pitch level by bilinguals

Level realization pattern of emotions is similar across the two languages: fear has the highest level, then it is followed by sadness and happiness, and anger has the lowest level. The main difference between the languages is that monolingual French children realize emotions with overlap, and neither of them separate significantly happiness from sadness, while monolingual Scottish children differentiate better their emotions, and all of them separate happiness from sadness. Bilingual children differentiate significantly their emotions by pitch level in their two languages. The significance of separation between happiness and sadness is no homogeneous across the bilinguals. Child 1 separates happiness and sadness in both languages (like Scottish monolingual children). Child 2 separates in English, and not in French (like monolingual children in the corresponding languages). Child 3 does not separate emotions in English, and separates happiness from the other emotions in French. Child 4 does not separate happiness and sadness in both languages (like French monolingual children).

### 5. Conclusions

A cross-linguistically comparable corpus of bilingual and monolingual children was recorded according to the developed methodology. In general, there is a lot of variability across all the children, but it is also possible to see some common tendencies in child affective realizations. The results show that bilingual children differentiate some emotions across their languages; happiness is the only emotion, which is differentiated by all the bilingual children. The majority of children (both bilingual and monolingual) realize differences between some emotions by pitch range measurements taken in this study. Monolingual children use analyzed acoustic parameters in a much more homogeneous way than bilinguals. Some results of bilingual children do not strictly correspond to those of monolinguals, and show bidirectional interference. Having a wider range of means and ways for affective realizations, bilinguals may represent a particular group of speakers who express vocal emotions in a different manner than monolinguals. Child data is currently compared to adult data (children's mothers) and used in cross-linguistic perception tests, developed to explore further the child mastering of affective prosody.

## 6. Acknowledgments

We express our gratitude to all the children and their parents who participated in this study. We are also pleased to thank Dr. H. Ouerdane and Dr. O. Gordeeva for their assistance with automatization of data extraction and processing.

## 7. References

- Scherer K. R., Banse R. and Wallbott H. G., "Emotion inferences from vocal expression correlate across languages and cultures.", Journal of Cross-Cultural Psychology, 32, 1, 2001.
- [2] Chen A. J., "Universal and language-specific perception of paralinguistic intonational meaning", PhD thesis. Utrecht: LOT. 2005.
- [3] Mejvaldova J., "Expressions prosodiques de certaines attitudes en tchèque et en français. Etude comparative.", PhD thesis, Univerzita Karlova, Université de Paris 7, 2001.
- [4] Aoyama K., Flege J. E., Guion S. G., Akahane-Yamada R. and Yamada T., "Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/." Journal of Phonetics, 32, 2004, pp. 333-250.
- [5] Mennen I., "Bi-directional interference in the intonation of Dutch speakers of Greek.", Journal of Phonetics, 32, 2004, pp. 543-563.
- [6] Ladd D. R., "Intonational phonology.", Cambridge University Press, 1996.
- [7] Patterson D., "A linguistic approach to pitch range modelling.", PhD thesis, University of Edinburgh, 2000.
- [8] Rose P., "Considerations in the normalization of the fundamental frequency of linguistic tone." Speech Communication, 6, 1987, pp. 343-351.