A methodological study into the linguistic dimensions of pitch range differences between German and English

Ineke Mennen*, Felix Schaeffler[†] and Gerard Docherty[‡]

*Centre for Research on Bilingualism, University of Wales, Bangor
†Speech Science Research Centre, Queen Margaret University, Edinburgh
‡ School of Education, Communication and Language Sciences, Newcastle University
imennen@qmu.ac.uk ; fschaeffler@qmu.ac.uk ; g.j.docherty@ncl.ac.uk

Abstract

The aim of this paper is to develop the methodology that would allow us to investigate the nature of variability in pitch range across speakers of different languages. In particular, we wanted to investigate whether by using linguistically based pitch range measures, such as those proposed by [1-3], we are able to characterise differences in pitch range across languages. We investigated Southern Standard British English (SSBE) and Northern Standard German (NSG), as it is often assumed that speakers of SSBE have a wider pitch range than speakers of NSG [4]. Using the linguistic measures suggested by [3], we found no such differences between NSG and SSBE, although a difference in the predicted direction *was* found with another linguistically based measure. Our study highlights the difficulty of using the previously suggested linguistic measures for cross-language comparisons, as some tonal structures are not equally distributed across the two languages. We therefore suggest that more suitable linguistic measures of pitch range may need to take the tonal distribution in the different languages into account.

1.   Introduction

It is often assumed that languages differ in their 'standard pitch range' [5;6]. This assumption is frequently based on comparisons of results from different studies with widely varying methodologies and potentially different factors influencing f0 (such as age, regional accent, physiology/anatomy, type of speech material). Very few systematic comparative studies of habitual pitch range differences between languages (or language varieties for that matter) have been carried out. Our study sets out to fill this gap by carrying out a large-scale systematic comparison of the pitch ranges employed by groups of speakers of Southern Standard British English (SSBE) and Northern Standard German (NSG). In this paper we will present results arising from a subset of our corpus.

1.1.   How to quantify pitch range?

The reason for the lack of systematic cross-language comparisons of pitch range may lie in the fact that pitch range is notoriously difficult to quantify successfully [7]. Although it is generally accepted that pitch range can vary along two dimensions, *level* and *span* [8], there is no consensus about the best way to quantify these, and many different measures of pitch range have been reported. Level (also referred to as register) reflects whether a speaker typically speaks with a relatively high or low pitch, whereas span reflects "whether a speaker's pitch covers a wide or narrow range of frequencies"

[3]. For level, measures of mean f0 and median f0 have been used. For span, the most commonly used measures are long term distributional (LTD) measures, including maximum minus minimum f0, four standard deviations around the mean, the difference between the 95th and 5th percentile (90% range), and the difference between the 90th and 10th percentile (80% range).

However, it has been suggested that there are some problems with using LTD measures since they assume an even distribution of f0 around the mean and are often affected by pitch tracking errors (e.g. octave errors) [3]. LTD measures were also shown to have rather weak correlations with listener judgements of speaker characteristics, suggesting that these measures lack perceptual validity [3].

An alternative to the classic measures of span and level was proposed by [3], building on work by [1] and [2]. The key to their measures of span and level is to link them to specific turning points in the f0 contour which, in turn, are linked to phonological tones and therefore thought to be linguistic in nature. We will refer to these measures as *linguistic measures*.

Patterson [3] showed that linguistic measures correlate better with listener judgments of speaker characteristics than the more commonly used LTD measures. The turning points he tested for the 'topline' of a speaker's pitch range included the average of phrase-initial accent peaks, non-phrase-initial accent peaks, and post-accent valleys. For the 'bottomline' he tested the average of phrase-final lows. His results showed that for *span* the best measure to capture variation across speakers appeared to be the average of non-phrase-initial accent peaks minus the average of post-accent valleys. The best measure for *level* appeared to be the average of phrase final lows.

1.2.   Quantifying pitch range across languages

The findings in [3] suggest that there is a strong basis for relating pitch range to averaged data taken from clearly defined linguistic targets in speech. However, although this method may give more perceptually valid results for characterising pitch range variation across speakers within a language, it is not clear whether it is a reliable method for cross-language comparisons. Furthermore, Patterson [3] points out that he based his linguistic measures of pitch range on the Bruce & Gårding [9] model of intonational analysis, rather than the more recent Pierrehumbert and Pierrehumbert & Beckman model [10;11]. An important difference between these two models is the fact that the latter assigns different phonological statuses to tonal targets, depending on whether they are associated to accented syllables or not. As a result of his choice of model, Patterson's measure of the 'post-accent

valley' does not distinguish between the different phonological statuses of L turning points. This distinction could be very important in cross-language comparisons where we expect to find a different distribution in the use of phonological tones (i.e. with a more frequent usage of L* in NSG as opposed to a more frequent use of H* in SSBE [12;13]).

Therefore, in our project as a whole we will (i) compare a wide range of both LTD and linguistic measures across 30 speakers of SSBE and NSG; (ii) investigate the contribution of the different statuses of phonological tones to our pitch range measures; and (iii) relate both types of measures to listeners' perceptual sensitivity to cross-language differences in production. Results on LTD measures in a subset of our corpus have been reported in [7]. In this paper, however, we will only present results on a range of linguistic measures in a subset of our corpus.

2. Method

2.1. Materials, subjects and procedure

We selected an English text [14] which was translated and slightly adapted for German. This text was read by a total of 60 speakers, 30 for each language. The subset of the corpus presented in this paper comprised a total of 22 speakers (i.e. 11 per language). The speakers were all female university students in their twenties and thirties and functionally monolingual (i.e. they were no more than moderately proficient in another language). The text was recorded together with materials recorded for another experiment not reported here. The speakers read the text once but were occasionally asked to repeat a sentence when they misread it.

The English text was recorded in a sound proofed room with a Marantz flash recorder and an AKG condenser microphone. The German recordings were performed under similar conditions with a Tascam DAT-recorder and an Audio-Technica condenser microphone. The test materials were digitised at 44.1 kHz sampling rate.

2.2. Measurements

Labelling was performed with Praat [15] according to the following procedure: The pitch contour was interpreted as a combination of pitch targets and linear interpolations between targets. Pitch targets (local maxima and minima) were derived from visual and auditory inspection. Additional pitch targets (changes in slope) were marked wherever interpolation between local maxima and minima did not lead to perceptually satisfying results when the original utterance was compared to an utterance with a resynthesised F0-contour based on the pitch targets. The changes in slope (marked as (D) and (U) in our material) were not included in the present analysis.

The local maxima and minima were then labelled as starred (H* or L*) or unstarred tones (H or L), depending on whether they were aligned with a stressed or an unstressed syllable. Separate labels were used for phrase-initial low, mid or high targets (IL, IM, IH), for phrase-final low, mid or high targets (FL, FM, FH) and phrase-initial accent peaks (H*i). No further assumptions were made about the status, association or combination of the tones. Figures 1 and 2 give examples of a labelled IP in both languages.
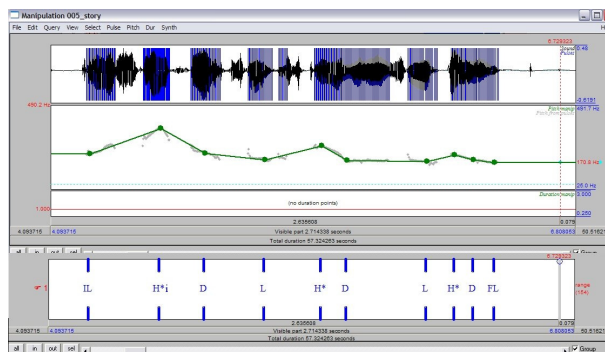

Figure 1: Pitch targets, labels and F0 contour for the English phrase "I was smoking outside the school when I saw Ben".
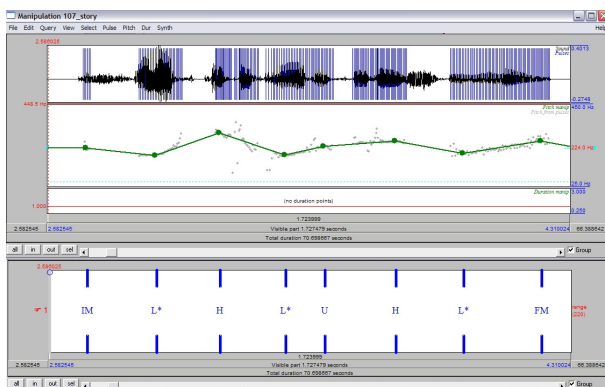

Figure 2: Pitch targets, labels and F0 contour for the German phrase "Ich rauchte gerade vor der Schule".

The following span measures were calculated: H-L, Hi*-L, Hi*-L*, Hi*-FL, H*-L, H*-L*, and H*-FL. The measures were derived by first calculating the speakers' averages for the tonal targets and then calculating the difference between the two respective targets in semitones (ST). For level, we used the mean FL values per speaker in Hertz.

3. Results

On inspection of the corpus it became obvious that our measure of 'non-phrase-initial accent peaks' (H*i) was not suitable for our cross-language comparisons, as there were hardly any accent peaks at the beginning of intonational phrases in our NSG corpus. This can also be observed in Figures 1 and 2. We therefore do not present the Hi* measures separately, but instead include them in our H* measures. That is, our H* measure represents the average of *all* accent peaks. Consequently, this could raise the top level of our span measures for SSBE, but see below for our results.

We first investigated the distribution of the different targets in our corpus. Figure 3 shows that there are considerable differences between the two groups in the distribution of tones. Whereas speakers of NSG more commonly use low accents (L*), H* are more frequent in SSBE. These differences in the distribution of tonal structures are also apparent in the post-accent peaks and valleys, which are predominantly H in NSG but more frequently L in SSBE.
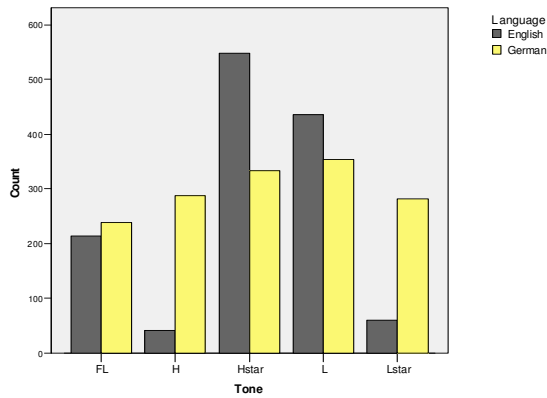
Figure 3: Distribution of measured targets in SSBE and NSG.

Not surprisingly, FL has a similar distribution across language groups.

As stated in section 1.2, contrary to [3], we distinguished in our measures between the different phonological statuses of L turning points. Figure 4 shows that the different statuses of these tones are indeed reflected in their f0 values. A mixed ANOVA with tone type (L vs. L*) as within-subjects factor and language as between-subjects factor showed a significant main effect of tone type [$F(1,20)=7.885$, $p=.011$]. The significantly lower values for the L* tones are in accordance with the predictions of the Pierrehumbert [10;11] model. The other factors (main effect of language and interaction between tone type and language) were not significant.

To test our main hypotheses we subsequently ran three separate t-tests with Bonferroni correction (significant $p<.016$) for each of the dependent span variables H-L, H*-L, H*-FL. As we expected wider span values for SSBE, directed hypotheses were formulated and one-tailed t-tests were used. For level, we ran a two-tailed t-test for the FL variable.

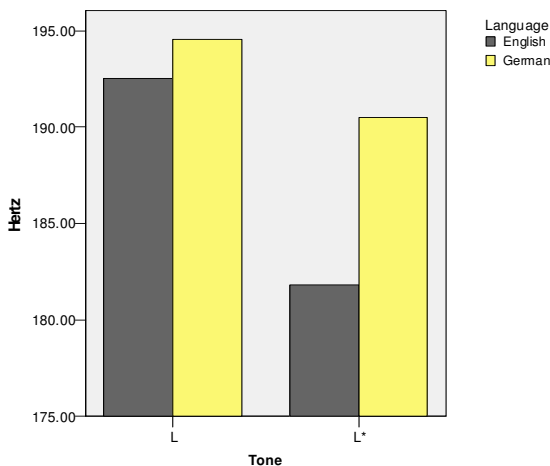Our results for span showed a significant effect of

Figure 4: Mean F0 values for L and L in SSBE and NSG.*

language for the measure H-L [$t(12.3) = 2.682$, $p=.01$, one-tailed t-test, equal variances not assumed], in the predicted direction (wider for SSBE). However, this effect was not

present in the H*-FL measure, nor in Patterson's [3] best span measure H*-L. We also did not find a significant effect of language in our FL measure for level.

Figure 5 is a scattergraph which represents one of the span (H-L in ST) and level (FL in Hertz) measures for all 22 speakers. From this figure it can be seen that speakers of NSG cluster at the lower end of the x-axis (representing span), while most speakers of SSBE cluster more at the higher end. It can also be seen that the speakers vary considerably in level, particularly in the group of SSBE speakers.

4. Discussion and conclusion

Only one of the four linguistic measures was successful in characterising pitch range differences across NSG and SSBE, in line with reported stereotypical beliefs and previously reported differences in long term distributional measures (from the same corpus, [7]) of a wider pitch span in SSBE. Thus the 'best' measure for our cross-language comparison appeared to be the difference between the average of post-accent peaks (H) and post-accent valleys (L). Interestingly, this is different from Patterson's [3] best measure, which was the difference between the average of non-initial-accent peaks and the post accent valleys.

We encountered various difficulties in using previously reported linguistic measures for our cross-language comparison. In particular, it was not possible to use the measure of phrase-initial accent peaks proposed by [3], as they hardly occurred in our NSG corpus. Patterson [3] distinguished between the phrase-initial and any following accent peaks because phrase-initial tone groups are known to be higher and have a wider pitch range than any following tone groups [16]. It is particularly this phrase-initial pitch
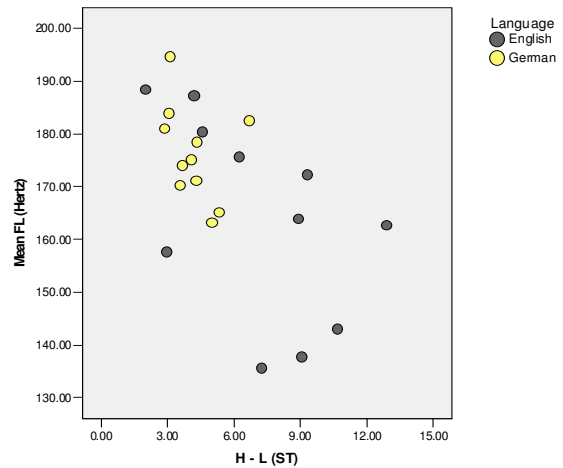

Figure 5: Scattergraph showing span (H-L) and level (FL) measures in NSG and SSBE.

accent which has been observed as being different in contrastive studies of German and English [17]. Perhaps the local pitch range differences of phrase-initial accents would have been a better measure to capture these differences. We intend to investigate this further in our future studies.

In conclusion, our findings suggest that it is not straightforward to apply previously established linguistic measures to cross-language comparisons. Even when

languages have a rather similar inventory of tonal categories, the distribution of these categories may be rather different, as observed in our corpus. Therefore, for linguistic measures to be useful for cross-language comparisons, they may need to take the tonal distribution in the different languages into account. It may, for example, turn out that listeners are not only sensitive to global pitch range differences, but may be influenced by the time that speakers spend near the top or bottom of that range. Future perception studies where various linguistic and long term distributional measurements are linked to listeners' perceptual sensitivity to cross-language differences may clarify these issues.

5. Acknowledgments

6. References

[1] Ladd, D. R., Terken, J. 1995. Modelling Intra- and Inter-Speaker Pitch Range Variation. *Proc.of ICPhS.* Stockholm, 386-389.

[2] Shriberg, E. and others. 1996. Modeling Pitch Range Variation Within and Across Speakers: Predicting F0 Targets When "Speaking Up". *Proceedings of the International Conference on Spoken Language Processing (ICSLP).* Philadelphia, PA.

[3] Patterson, D. 2000. *A Linguistic Approach to Pitch Range Modelling.* PhD dissertation, University of Edinburgh.

[4] Eckert, H., Laver, J. 1994. *Menschen und ihre Stimmen: Aspekte der vokalen Kommunikation.* Weinheim: Psychologie Verlags Union.

[5] Braun, A. 1994. Sprechstimmlage Und Muttersprache. *Zeitschrift Für Dialektologie Und Linguistik* 61(2), 170-178.

[6] Chen, A., Gussenhoven, C., Rietveld, T. 2004. Language-Specificity in the Perception of Paralinguistic Intonational Meaning. *Lang. Speech* 47, 311-349.

[7] Mennen, I., Schaeffler, F., Docherty, G. J. 2007. Pitching It Differently: a Comparison of the Pitch Ranges of German and English Speakers. *Proceedings of the 16th International Congress of Phonetic Sciences.* Saarbrücken, Germany.

[8] Ladd, D. R. 1996. *Intonational Phonology.* Cambridge: Cambridge University Press.

[9] Bruce, G., Gårding, E. 1978. A Prosodic Typology for Swedish Dialects. In: Gårding, E., Bruce, G., Bannert, R. (eds), *Nordic Prosody* Gleerup, 219-228.

[10] Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation.* PhD Dissertation, Cambridge, MA: MIT.

[11] Pierrehumbert, J., Beckman, M. 1988. *Japanese tone structure.* Cambridge, MA: MIT Press.

[12] Anderson, K. O. 1979. On the Contrastive Phonetics of English and German Intonation. Festschrift Für Otto Von Essen Anlässlich Seines 80. Geburtstages. In: Wängler, H.-H. (ed), *Hamburger Phonetische Beiträge 25* 25-35.

[13] Delattre, P. 1965. *Comparing the phonetic features of English, German, Spanish and French.* Heidelberg: Julius Groos.

[14] Brown, A., Docherty, G. J. 1995. Phonetic Variation in Dysarthric Speech As a Function of Sampling Task. *Eur. J. Disorder. Comm.* 30(1), 17-35.

[15] Boersma, P., Weenink, D. 2006. Praat (Version 4.5). http://www.praat.org.

[16] Clark, R. A. J. 1999. Using Prosodic Structure to Improve Pitch Range Variation in Text to Speech Synthesis. *Proceedings of the 14th International Congress of Phonetic Sciences.* San Francisco.

[17] Trim, J. L. M. 1988. Some Contrastive Intonated Features of British English and German. In: Klegraf, J., Nehls, D. (eds), *Essays on the English Language and Applied Linguistics on the Occasion of Gerhard Nickel's 60th Birthday.,* Heidelberg: Julius Groos, 235-249.