Cues for Perceived Pitch Register

Toni Rietveld and Patricia Vermillion

Department of Language and Speech
University of Nijmegen, The Netherlands
A.Rietveld@let.kun.nl

Department of Linguistics
QMW, University of London, England
P.Vermillion@let.kun.nl

Abstract

The aim of this experiment was to assess empirically listeners’ behaviours in characterising pitch contours with the label pitch register. The motivation for this assessment was initiated by the conflicting use of the term ‘register’ in speech science and intonology. The findings reported here indicate that pitch register would more appropriately be associated with position of the Low pitch targets and the mean value of the tonal contour. In addition, the importance of the F0-min and the distance between H and L targets have been found to be weaker than previously assumed.

1. Introduction

The term register has different meanings in different fields. In speech pathology or speech science it is essentially the same as vocal mode, which is a term used on a discrete scale to describe laryngeal behaviour [8]. The mode or register permitting highest pitch is called falsetto, and the mode with the lowest pitch is vocal fry. In intonation, register has a different meaning, and it is commonly used on a continuous scale. Three descriptive labels for overall characteristics of pitch contours are used [3,5,7]: Reference Frequency (Fr or F0-min), Range and Register, which are illustrated in figure 1.

![Figure 1 Range and Register according to [3, 5, 7].](image)

In papers written in the field of intonation, the concept of range is used to express the distance between the lowest F0 value (‘F0-min’) reached by the speaker (often estimated by L%) and the location of H targets. F0-min is a speaker-specific value, which appears not to be affected by raising or lowering the voice [4, 6]. The label register is used to parameterise the distance between targets of H and L. The expression used to model F0-variation [1,5] is given in (1):

\[ F0 = Fr \times N^{W \times T} \]  

(1)

(without provision for Downstep factors) with \( T = +1 \) for high tones and \(-1\) for low tones.

The model defines high and low grid lines; increasing \( N \) causes both the high and low grid-lines to be raised by some multiplicative factor, with a stronger effect on the H than on the L grid-lines (thus \( N \) has, in effect, a more direct relationship with the concept of ‘range’ than with ‘register’). Increasing \( W \) will raise the H, and lower the L grid-lines (and is thus more directly associated with ‘register’).

In one perception experiment, it was found that increasing \( Fr \), \( N \) and \( W \) positively affects attributes like ‘liveliness’ [3]; the effect of \( Fr \) and \( N \) largely exceeded that of \( W \). When we accept that subjects will base their assignment of ‘liveliness’ to the degree of variation between H and L (distance between H and L), \( N \) primarily accounts for register.

As F0-min will only seldomly be reached, we hypothesize here that the perceived range is the distance between H and L targets, as generally more samples of H and L will be available to the listener than samples of L%. Unfortunately this ‘perceived range’ already has a label in intonology: ‘register’. We believe that the term ‘register’ should be used to label speaker’s behaviour as ‘speaking up’ or ‘speaking down’. A prerequisite for the use of this term is the ability of naive listeners to label speech on scales like ‘register’ and ‘range’. For example, reporters who give a live-report of a horse race tend to raise their voice when the horses approach the finish line in comparison to the beginning of the race. This is especially evident if the horses are close together and the outcome of the race cannot be predicted. Although the reporter’s voice rises, s/he does not necessarily increase the distance between his/her H and L targets. We assume that listeners will tend to use the term ‘register’ when describing this vocal behaviour.

The question still is, which characteristics of the pitch contour give rise to the perception of high or low register. However, a more fundamental question is whether we need perceptual correlates of the concepts ‘register’ and ‘range’ which can be used by (naive) listeners. The concept of ‘register’, as illustrated in Fig. 1, has proven to be a fruitful concept, as it enables intonologists to parameterise tonal phenomena like ‘downstep’, and ‘accent’. In the field of intonology the name of the label does not seem to be of any relevance. In the field of the paralinguistic and cross-linguistic use of intonation, however, the name of the label does matter, as listeners do play a role in the experimental paradigms associated with this type of research. One hypothesis, namely the frequency code, states that high values of F0 tend to code ‘submissiveness’ [12]. A second hypothesis asserts that high F0 values also tend to code attributes like ‘emphasis’ [13]. In the context of these theories, it is important to know whether a speaker...
realises his/her utterances in a ‘high’ or in a ‘low’ register as local tonal phenomena can get a different meaning in the context of perceived register.

Listeners are well able to characterize pitch contours in terms of pitch register and pitch range, as was demonstrated by [2]. Moreover it was found clear predictors for both attributes. For register it was median F0 (PM correlation = 0.80) and for range the variation coefficient of F0 (s.d./mean): PM correlation = 0.68. [2] used recorded, non-manipulated speech, and didn’t report any tonal specifications. We don’t know, for instance, whether the contours consisted of H* or L*-accents, whether the boundary tones were %L, %H, L% or H%, and whether the contours were of the linked type (for instance H* +L, H*). In the former there are more high F0-values than in the latter.

We do not know at this point, which cues are used and which weights are assigned to them. In order to explore the role these cues may play, we carried out the following perception experiment. In this experiment, we test the validity of an alternative use of the concepts ‘register’ and ‘range’, and assessed the weights attributed by listeners when rating pitch contours on a ‘pitch register’ scale. We hypothesize the following: (1) Listeners rely more heavily on H and L targets than F0 min when determining pitch register. In light of this hypothesis, we also predict that (2) Listeners use values of tonal targets, rather than the mere distance between H and L values when determining pitch register. Finally, we expect that (3) Listeners exploit the mean value of a tonal contour more than the mid point.

2. Methodology

The following experiment addresses the questions of which cues are used in the perception of pitch register and how much weight is assigned to them. In order to explore the role these cues may play, we have adopted the following methodology, which assesses listeners’ perception of perceived pitch register.

2.1. Stimuli

The source data used in this study was taken from a previous study conducted by the second author [14]. An adult male speaker of RP British English was instructed to read aloud from a provided script, “as naturally as possible”, whilst being recorded onto a Sony 90-P digital audio tape (DAT) using an Aiwa HD-S 100 DAT recorder and a Sony ECM-MS 957 microphone. A set of 72 stimuli was derived from the utterance: Earl Grey tea is not good for you.

The source utterance was stylised and seven pitch points were located at tonal movement positions, which are given as times in Figure 2.

The Total Rescaling method [10], or an up/down-scaling of all pitch points within the contour, was adopted. All pitch points were adjusted to one of eight pitch levels (PL), which were derived from the average group range from thirty male (98-136Hz) and female (188-238 Hz) British speakers [11]. The Hertz values were converted into Semitones (ST), initiating the highest and lowest PLs for men and women as 0-3.5 ST and 11.1-14.5 ST, respectively. Two steps equal in ST value were found between the maxima and minima PLs, which were then adopted as the third and fourth middle PLs. Specifically, the PLs (in ST) selected for the male and female speakers included (1) 0 (2) 1.17 (3) 2.34 (4) 3.5, and (1) 11.1 (2) 12.23 (3) 13.36 (4) 14.5, respectively. Two anchor stimuli, derived from the same source utterance, served as the comparable, “neutral” register. Both anchors were adjusted using the TR method to one of two average F0s, selected from the mid-point of the female range (12.8 ST) and male range (1.75 ST).

Using the seven pitch points, four types of manipulations were done for each of the eight assessed PLs. First, three contour types were created in order to determine whether listeners use mean value or mid point when assessing pitch register. Unique only in mean score, the three contours are illustrated in Figure 3.

2.2.1. Stimuli

The speech material was digitised with a sampling rate of 48 kHz, which was later downsampled to 16 kHz in mono. With the help of a signal processing system, namely Praat [9],
In order to determine whether listeners use L values when ascertaining pitch register, all of the L values were adjusted to one of two heights, including a high L value (PP 1.3, 4.6 = PL - 1 ST; PP2.5 = PL + 2.5 ST) and a low L value (PP 1.3, 4.6 = PL - 8 ST; PP2.5 = PL + 2.5 ST). Figure 5 depicts the manipulation of the L values.

![Figure 5. Low L values (---) and High L values (•••)](image)

Finally, we examined whether listeners employ the L% when determining pitch register; the F0 minus was manipulated creating a low L% by subtracting 5 semitones from the previously fixed L% value (1.1 for women and -10 for men) or adding 5 ST for the high L%. Figure 6 illustrates the manipulation of L%.

![Figure 6. Low L% (---) and High L% (•••)](image)

Collectively, the manipulations created 72 stimuli: (2 widths) + (2 L values) + (2 L%) + (3 mean values) + (8 pitch levels).

Each stimulus was concatenated with an anchor utterance. The resulting stimuli were low-pass filtered using Cool Edit, with a cut-off frequency of 110 Hz, and a filter order of 3.5. The utterances were then amplified 350 per cent. Finally, the resulting stimuli were assessed for lexical comprehension; the utterances were played randomly for four individuals, none of which could distinguish the lexical make-up of the utterance.

The stimuli were then randomised four times, which were recorded onto DAT and later transferred to four separate Maxell Minidisks using a Sony MXD-D5C Minidisk Recorder and a Sony TCD-D8 Recording DAT Walkman.

### 2.2. Procedure

In total, 30 university students at the University of London, Queen Mary and Westfield College took part in this experiment; all were native English speakers. Three separate groups of participants were first instructed on the meaning of pitch register by providing a verbal explanation, in addition to examples illustrating pitch register. The two examples provided included a person speaking to an infant (higher pitch register), and a bored student (lower pitch register). The stimuli were presented over a Yoko Pro-400 speaker, using a Sony MZ-R55 minidisk recorder/player.

The participants were instructed to rate the second utterance (test stimuli) in relation to how much higher or lower the pitch register was from the first utterance (anchor stimuli). The judgements were obtained using a magnitude estimation scale, measuring 100 millimetres in length. The data were measured by the second author, according to a 100-millimetre scale.

### 3. Results

Separate analyses of variance (repeated measures, alpha-level 0.05) were carried out on subsets of the data, always with one between-subject variable: sex of listeners.

- **Subset I:** Three within-subject factors: Mean value with three levels: High, Low, and Mid. Sex of speakers (2 levels) and Pitch level (4 levels).
- **Subset II:** Three within-subject factors: Width (2 levels: wide and narrow), Sex of speakers and Pitch level.
- **Subset III:** Three within-subject factors: Level of L (2 levels: relatively high and relatively low), Sex of speakers and Pitch level.
- **Subset IV:** Three within-subject factors: Level of F0 minus (2 levels: relatively high and relatively low), Sex of speakers and Pitch level.

Apart from the omni-present effects of ‘Sex of speaker’ and ‘Pitch level’, the following significant effects were found in the four subsets of data; they are presented in Table I, together with $\eta^2$ as measure of effect size. In order to compare the magnitudes of the effects, we also present $\eta^2$ associated with the factor pitch level in each dataset; the factor pitch level is the most global, and physically realistic measure of pitch register.

**Table 1. Significant effects (at 0.05 level) and associated values of $\eta^2$ together with $\eta^2$ associated with factor Pitch level in each of the four data sets.**

<table>
<thead>
<tr>
<th>Effect</th>
<th>$\eta^2$ Pitch level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Mean (rel. high,medial,low)</td>
<td>0.203 0.219</td>
</tr>
<tr>
<td>2a) Width (narrow, wide)</td>
<td>0.029 0.202</td>
</tr>
<tr>
<td>2b) Sex of Speakers × Pitch level</td>
<td>0.018</td>
</tr>
<tr>
<td>3a) Lows (rel. high, low)</td>
<td>0.382 0.143</td>
</tr>
<tr>
<td>3b) Lows × Pitch level</td>
<td>0.011</td>
</tr>
<tr>
<td>4a) F0 min (rel. high, low)</td>
<td>0.026 0.324</td>
</tr>
<tr>
<td>4b) Sex listeners × Sex speakers × Pitch level</td>
<td>0.003</td>
</tr>
</tbody>
</table>

All factors implemented in our design, including Mean F0 value, Width (distance between H and L), relative position of L values and relative position of L% (F0-min) turned out to be significant. Just a very small number of significant interactions with extremely small effects sizes were found; they can be left out of discussion. In terms of effect size, the position of the Low targets was the most important ($\eta^2 = 0.382$), with a rating of 67.1 and 69.4 for high L values and 43.4 and 39.2 for low L values, for male and female speakers respectively.
The values of Mean F0 were the second largest effect size ($\eta^2 = 0.203$). Comparing the high, medial and low Mean scores, the ratings were 67.7, 6.01 and 52.0 for male speakers and 7.15, 63.0 and 5.10 for female speakers, respectively.

Both the effect sizes of ‘Width’ (the distance between H and L, labelled by [3, 5, 7] as ‘Register’), and of F0-min are very small ($\eta^2 = 0.157$), although the raw mean scores show effects. Specifically, the raw means for the F0-min are 53.7 and 51 for the high values and 50.1 and 45 for the low values for male and female speakers respectively. For pitch width, the pitch register ratings for the narrow contours were 56.9 and 51 for the high values and 50.1 and 45 for the low values for male and female speakers respectively. For pitch width, the pitch register ratings for the narrow contours were 56.9 and 51 for the high values and 50.1 and 45 for the low values for male and female speakers respectively.

4. Conclusions

In conclusion, the assessment of a listener’s judgment of pitch register according to mean values or mid points, the mean value of a tonal contour was confirmed to function as a cue for pitch register, which confirms [2]. Second, the relative level of the L targets was the strongest cue indicating pitch register. This finding infers that H and L targets are, in fact, used by the listener to determine pitch register, as hypothesised in this paper. In addition, this result also supports the finding that listeners use the mean value when determining pitch register, as increasing the L values affects the overall mean of the tonal contour. Third, the effect of F0-min was relatively small, which confirms our original hypothesis that listeners need more information than the final tonal cue. Foremost, the effect of pitch width was smaller than expected in consideration of the position it takes in previous attributions made in intonology. In addition, the narrow contours were rated as having a higher pitch register than the wide contours. This unexpected response behaviour may be due to L values being higher in the narrow range than the wide range. Therefore, this finding supports the previous evidence that listeners use L values when determining pitch register.

5. References


Acknowledgements

We are grateful for the assistance provided by Joop Kerkhoff in various stages of this experiment. We would also like to acknowledge Carlos Gussenhoven and Aoju Chen for their useful comments regarding the design of the experiment.