

Intrinsic pitch differences between German vowels /i:/, /ɪ/ and /y:/ in a cross-linguistic perception experiment

Daniel Pape¹, Christine Mooshammer², Susanne Fuchs¹ and Phil Hoole³

¹ ZAS Berlin, Germany

² IPDS (University of Kiel), Germany

³ IPSK (LMU Munich), Germany

pape@zas.gwz-berlin.de

Abstract

Perceived pitch differences between high vs. low vowels with identical F0 have been reported in the literature. On the speech production side, in German a specific phenomenon is found: Tense and lax vowels differ in their tongue height but tend to have a similar F0. On the speech perception side, our aim was to test whether German listeners judge the tongue height or real F0 of tense and lax vowel pairs. Therefore a vowel pitch comparison experiment was conducted for three different German vowels /i:/ y:/ ɪ/ showing a similar phonological vowel height, but differing in their tense/lax distinction and roundedness. The second question examined was whether pitch perception was influenced by language inventory. Therefore the same perception experiment was conducted cross-linguistically with Catalan listeners in Spain. Results show that the difference in tenseness was not significant in either language family, indicating that listeners judged the physical F0 instead of underlying articulatory settings. Roundedness affected pitch differences in both languages. Musical education influenced sensitivity to pitch differences only for German listeners. Surprisingly, Catalan listeners were insensitive to even large F0 changes and judged the pitch differences in vowels mainly based on the categorical vowel identity, independently of their musical education.

1. Introduction

When the pitch of different vowels is compared, it was consistently found that low vowels, i.e. /a:/, are perceived higher in pitch than high vowels, i.e. /i:/, when presented with the same nominal fundamental frequency (F0) [1] [2]. This perceptual phenomenon is called “intrinsic pitch” of vowels (IP) and also applies to German diphthongs [3]. However, explanations of the observed pitch difference are inconsistent:

- Fowler and Brown [2] who found a perceived pitch difference of 3.4Hz between high and low vowels of the same F0, assumed that this difference would be a compensation for the well-known “intrinsic F0” (IF0) phenomenon. It consistently describes a speech production difference in the fundamental frequency of about 10-15 Hz between high and low vowels in nearly all languages [3]. This would lead to a different prosodic contour comparing high and low vowels for the same target. Following Fowler, IP is an abstract compensatory mechanism which compensates IF0 at a perceptual level in order to yield a consistent speech melody, independent of the identity of the target vowel, and therefore aims to facilitate “perceptual parsing”. However, no explanation for the different frequency amount in perception

(about 1/10th of the production difference) compared to production was provided, which contradicts (at least a complete) compensation.

- Stoll [5] claimed purely psychoacoustic reasons for the different perceived pitch of high and low vowels which is in contrast to the theory of Fowler. Due to the different spectra of the vowels a small but perceptually significant shift of the perceived pitch is introduced, according to the “virtual pitch shift theory” of Terhardt [6]: He explains the perceived pitch of a complex sound as the average of a weighted network from pitch impressions by harmonics and their sum and difference tones, which can introduce a small shift of the perceived pitch.

- In languages like German and English the vowel space is rather crowded (i.e. 15 unreduced vowels for German). Therefore it could be possible that the pitch of a vowel is used as a perceptual cue to simplify vowel identification and to avoid perceptual confusion. Evidence for this theory was given in a study by Traunmüller [7]. He found that the frequency distance between fundamental frequency and first formant is used perceptually to define the “openness” category of a vowel. However, in other studies [8] no dependency of the fundamental frequency on the identification of vowels has been found.

For the tenseness category in German it has been found that in speech production tense and lax vowels are produced with about the same F0 although their tongue height differs [9] [10]. However, according to a biomechanic explanation for IF0 [11] the F0 for tense vowels should be higher compared to their lax counterparts. An explanation for the similar F0 for the tense and lax vowel pairs was given by experiments of Hoole et al. [12] who found a higher CT muscle activity for lax vowels, indicating an active F0 raising for these vowels in speech production. Examining speech perception, the question arises whether tense vowels evoke a different pitch compared to lax vowels when presented with the same F0. Thus the first aim of the current study was to examine if pitch perception differences between tense and lax vowels can be found: If these differences exist then listeners would judge the underlying articulatory settings, if no differences can be detected then listeners would judge the real F0 of the vowels. The second aim was to examine the pitch perception of vowels with different frequency damping due to radiation differences between rounded and unrounded vowels by means of the inclusion of a rounded vowel with a similar vowel height. It allows to test the hypothesis of Stoll, who claims that psychoacoustic pitch shift causes IP differences, since in our experiment these differences are eliminated.

The perception experiment was designed to compare German with a language which neither shows a tenseness nor a roundedness distinction for the vowels in question. Thus the

third aim of the current study was to examine how listeners of such a language would behave when merging multiple vowel categories (like the pitch perception of vowels of a similar vowel height) into a single vowel category. Catalan was chosen to examine this question.

2. Method

2.1. Stimulus preparation

The vowels /i:/, /y:/ and /ɪ/ were chosen because they are similar in vowel height and intrinsic fundamental frequency [9] [10]. They differ phonetically in tenseness: tense for /i:/ and /y:/ and lax for /ɪ/. Further, due to the roundedness of /y:/ the spectral structure and slope between /i:/ and /y:/ is different, with significant higher energy in higher frequencies for /i:/ and /y:/ than for /ɪ/. The stimuli were cut from a high-quality recording of a standard German speaker in a sound-treated room (Microphone Sennheiser MKH20 recorded onto DAT).

The vowels were embedded in the bilabial stop context /bVpə/ and were extracted from the burst of the preceding plosive to the burst of the following plosive. Since the target word was stressed the F0 contours showed a rising slope with a range of 25Hz. The duration was 250ms for /i:/, /y:/ and 125ms for /ɪ/ which is the standard length for these vowels found in literature. The stimuli were selected in such a way that they show the same F0 contours with the same starting and end F0 values and a similar slope. Table 1 gives the formant values. Obviously due to the shorter phonetic duration of the lax vowel /ɪ/ the slope is steeper compared to the tense counterparts.

With the algorithm of Terhardt [6] the virtual pitch was calculated for the stimuli. The values are given in table 1 and indicate that the virtual pitch shift values are identical. Therefore it can be excluded that this psychoacoustic phenomenon could explain any pitch biases measured in the experiment.

Table 1: Formant values and F2' (perceptual substitute for higher formants, see section 3.1.) and virtual pitch values (according to the model of Terhardt [6]) for the stimuli of the perception experiment.

	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]	F2' [Hz]	Virtual Pitch [Hz]
/i:/	301	1988	2854	3151	2664	133.1
/ɪ/	389	1625	2298	3272	2398	133.1
/y:/	359	1583	1878	3211	2224	133.5

Since pitch perception is dependent on the loudness of the presented stimulus [13] [14], the stimuli were adjusted to equal loudness. A study of the World Broadcasting Union [14] compared different loudness measurements by correlating them to the results of an exhaustive perception experiment and found that a simple RMS (SPL_{LEQ}) measurement gave the best correlation to the human perception of loudness and was even superior to complicated psychoacoustic models. Therefore by applying level amplification the stimuli were adjusted to the same RMS (SPL_{LEQ}) values.

Finally, the processed vowel prototypes were pitch-shifted with the PSOLA algorithm with the standard settings of the software PRAAT [15]. The range was ± 10 Hz in 2.5Hz

steps (corresponding to about 2 semitones at the given F0). The chosen methodology for the difference perception was the 2I2AFC test: Listeners were forced to judge if the first or the second stimulus in a given pair was higher in pitch. If uncertainties occurred, the listener was allowed to repeat the pair before making the judgement. A set with 10 stimuli were presented beforehand to practice the procedure. Three sets with 70 stimuli pairs were run. Stimuli were paired in randomized order.

2.2. Listeners

The perception experiment was carried out (1) in Berlin (Germany) with 25 native German listeners and (2) in Tarragona (Spain) with 32 native Catalan listeners. Only Catalan listeners were chosen not educated in Germanic languages English or German. All listeners were asked about their musical education: 16 of the German listeners and 14 of the Catalan listeners reported musical education. All listeners reported normal hearing.

Additionally to written instructions, the investigator explained the procedure to the listeners by speaking different vowels at different pitches to make sure they understood the task. Furthermore, for musically educated listeners, the sentence: "The important task is the difference in pitch, for example playing different notes on a piano, NOT the colour of the sound, like playing the same note i.e. "b-flat" either on a piano or on a violin" was given to ensure that listeners would judge the pitch and not the timbre of the vowels. Following the practice set, listeners were asked if they had any uncertainties or questions.

3. Results

3.1. Differences in pitch perception to sound equal in pitch

Following Fowler and Brown we applied PROBIT analysis to fit ogives to the curves of individual subjects. Our dependent variable was the F0 difference between the vowels to be examined at which, on the fitted ogive, subjects judged the tense vowel higher on 50% of the cases. Because the response patterns of 8 German and 24 Catalan subjects were markedly nonmonotonic across the continuum, they were treated differently (see section 3.2.).

For the German listeners, this described value averaged 0.9Hz for the /i:/-/ɪ/ distinction and -4.2Hz for the /i:/-/y:/ distinction. Only for /i:/-/y:/ was a significant departure from 0Hz found ($T(11)=-4.86$, $p<0.001$). That means that to sound equal in pitch, the vowel /y:/ has to be presented with a higher F0 of 4.2Hz in reference to the F0 of /i:/.

For the Catalan listeners values of 3.2Hz on average for the /i:/-/ɪ/ distinction and -10.8Hz for the /i:/-/y:/ distinction were observed. Again, only for /i:/-/y:/ was a significant departure from 0Hz found ($T(8)=-3.376$, $p<0.012$).

Figure 1 shows means for all German listeners in the upper panel and Catalan listeners in the lower panel (note that this figure shows means over all listeners, therefore a difference to the above reported numerical values occurs). As was also found by Stoll [5], spectral shape, i.e. lip rounding, has a significant effect on pitch judgements, with significant lower pitch for the rounded vowel compared to the unrounded vowel. This effect was significant in both languages, with an

even lower perceived pitch in the Romance language. Contrary to Stoll's theory, in our data "virtual pitch shift" could not explain this significant difference in pitch perception.

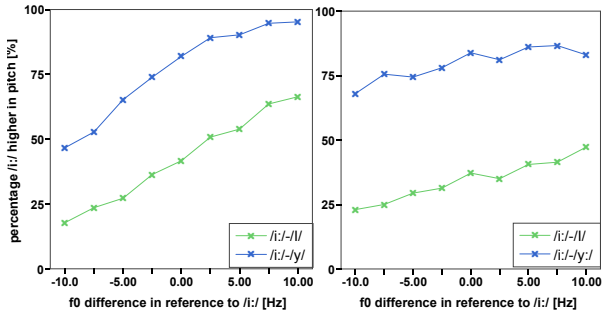


Figure 1: Mean values for the judgement „i:/ is higher in pitch“ in dependence on the physical F0 difference. In dark colour the responses for /i:/-/ɪ/ and in light colour the responses for /i:/-/y:/: are shown (upper panel German results, lower panel Catalan listeners).

This influence of spectral shape on pitch judgements could be explained by the phenomenon of "sibilant pitch" [16], which is the perceived pitch of whispered vowels: Traunmüller [16] found that this pitch (which occurs in absence of glottal vibrations) corresponds to F2', an average of the higher formants starting from F2. Furthermore, Carlson [17] found 3210Hz for the F2' in Swedish /i:/, compared to /y:/ with 2010Hz (when F0 and F1 were fixed).

Thus, listeners asked to judge the pitch of a sound mainly judge "fundamental pitch", but are strongly influenced by a pitch perception evoked by spectral energy allocation, introducing a bias in the pitch comparison. Since in a pitch perception experiment it is impossible to be sure if listeners' judgements are biased to a greater degree by fundamental pitch or by sibilant pitch, an interference of this factor cannot be excluded. In table 1 the F2' values for the stimuli are given. As can be seen, sibilant pitch for /y:/ is lowest with 2224Hz. Therefore the difference to the sibilant pitch of /i:/ (2664Hz) could explain the perceived significant pitch difference for both Catalan and German listeners (note that the vowel quality for Swedish /y: i:/ is different compared to the German vowel quality for these vowels, therefore differences in sibilant pitch values could occur).

However, it could be expected that sibilant pitch differences between /i:/-/ɪ/, (2664Hz and 2398Hz, respectively) could also introduce a significant pitch difference, at least for the Catalan listeners, who do not have a linguistic distinction between these vowels and therefore should judge the differences based on acoustic parameters. But no significant pitch difference for these vowels can be found. Therefore it is assumed that the sibilant pitch difference between these vowels (which is 220Hz in comparison to 440Hz for /i:/-/y:/) is too small to introduce a significant bias.

3.2. Accuracy of the response function towards F0 changes

In order to examine the influence of the native language on the reliability of the pitch judgments the deviation of the listeners pitch judgment was compared to a "linear judgement". This linearity was defined as a linear rise from 0% "higher in pitch" at -10Hz difference to 100% at 10Hz

difference with a 50% "by chance" judgements at equal F0s. This function would result from an experiment with a "perfect listener" in a condition in which a comparison of simple complex tones is made. For each "real" listener the sum of each difference value of his/her pitch judgements from the reference values were computed (variable diff), indicating the accuracy of pitch responses to the given stimuli. Therefore small values indicate pitch judgements close to the ideal response, whereas large values indicate an insensitivity to the F0 differences (categorical behavior) or large pitch bias (parallel shift to the ideal function).

Table 2 gives the results of a significance test for the variable "diff", the values is significantly different for the two languages with smaller values for German. Differences between the languages were significant for each vowel comparison and when aggregated over both vowels.

Table 2: t-values and significance levels for the difference of the accuracy of pitch judgments with reference to an ideal response (see text) depending on the listeners native language. Significant values are printed in bold.

All vowels	t(54)=3.272	p< 0.002 **
/i:/-/ɪ/	t(54)=3.092	p= .0003 **
/i:/-/y:/	t(54)=2.464	p= 0.0017 **

Table 3 gives the F-values and significance levels for the accuracy of pitch judgement for each language depending on musical education, which in the literature is often found to improve pitch discrimination accuracy (see i.e. [18]). The dependent variable was the deviation value for each F0 difference, split by language. As can be seen, in German musical education has a significant effect on the accuracy of the pitch responses for the /i:/-/ɪ/ pair, the /i:/-/y:/ pair and the aggregation over both vowels. For the Catalan listeners no significant effect of musical education could be detected.

Table 3: t-values and significance levels for the difference of the accuracy of listeners pitch judgements on musical education. Significant values are printed in bold.

	German		Catalan	
All	t(430)=4.0	p= .000 **	t(574)=1.38	p= .168
/i:/-/ɪ/	t(214)=3.2=	p= .014 *	t(286)=.852	p=.395
/i:/-/y:/	t(214)=2.5	p= .002 *	t(286)=1.1	p=.273

4. Discussion and Outlook

The presented study examined the existence of pitch differences for German vowels with a similar phonological vowel height but differing in tenseness and roundedness. The study was conducted with (1) listeners of a Germanic language and (2) listeners of a Romance language to examine possible cross-linguistic effects. The results indicate that observed differences could not be due to psychoacoustic pitch shift which is considered as one of the possible reasons for intrinsic pitch differences [5]. The differences in tenseness were not significant in either language family, therefore pitch does not seem to be perceptually exploited by listeners to enhance vowel differentiation for the tense-lax distinction, at least for the conditions in the given experiment (vowels in stressed position in their original phonetic length in stop context). The possibility remains that the pitch of the vowels is used to enhance vowel identification when the vowels are spectrally very similar, i.e. to identify reliably /ɪ/ vs. /e:/. To

examine this question other pitch experiments in combination with vowel identification will be conducted.

Furthermore, results indicate a systematic difference in pitch perception and discrimination for the Germanic language compared to the Romance language when examining pitch discrimination for German vowels: Catalan listeners show an insensitivity to even large F0 differences and judge the pitch differences in vowels mainly based on vowel identity. Surprisingly, this insensitivity in pitch discrimination of Catalan listeners was not dependent on their musical education. It is not clear what mechanism causes the existing different pitch sensitivity in the two languages: It should be expected that musical education improves pitch sensitivity and improves therefore the sensitivity to pitch differences. However, most musically educated Catalan listeners showed an insensitivity to pitch differences of up to 2 semitones. Generally, it should be expected that Catalan listeners would be more sensitive to the pitch differences in the German vowels, since these vowels are not included in their native vowel system and should therefore be judged more accurately by the listeners.

To our knowledge pitch perception of vowels was only investigated in Germanic languages. A follow-up experiment will test if a universality of IP between high and low vowels exists since our results indicate a different pitch discrimination in Romance languages, at least for Catalan. Result from research in neuroscience indicate that it is not clear if separate pitch perception mechanisms for speech melody and music melody exist. On the one hand researchers, i.e. Patel [19], provide evidence for different processing for speech prosody and musical melody by means of neuroimaging. On the other hand researchers, i.e. Schon et al. [20], show results which speak for a shared processing for prosodic and melodic structure. It would be interesting to examine whether pitch discrimination differences exist for vowels on the one hand and musical tones on the other hand. To the authors knowledge such a comparison has not been published yet. The results of such an experiment will shed light to the pitch processing of speech and musical tones.

5. Acknowledgements

This work was supported by a grant from the German Research Council GWZ 4/10-1, P.1. We would like to thank Joaquín Romero and Sydney Martin for the professional and amazingly kind collaboration during the realisation of the perception experiment in Tarragona (Spain). Also we are thankful to B.C.Moore, H.Traunmüller, W.Serniclaes and J.Brunner for helpful comments.

6. REFERENCES

[1] Chuang K. and Wang W.S.Y. "Psychophysical pitch biases related to vowel quality, intensity difference and sequential order", *Journal of the Acoustical Society of America*, vol.64, no.4, pp. 1004-1014, 1978.

[2] Fowler C.A. and Brown J.M. "Intrinsic f0 differences in spoken and sung vowels and their perception by listeners", *Perception and Psychophysics*, vol.59, no.5, pp.729-738, 1997.

[3] Niebuhr O. "Intrinsic pitch in opening and closing diphthongs of German", in B.Bel, I. Marlien (Eds.) *Proc.*

International Conference: Speech Prosody 2004, Nara, Japan, 2004.

[4] Whalen D.H. and Levitt A.G. "The universality of intrinsic F0 of vowels", *Journal of Phonetics*, vol.23, 349-366., 1995.

[5] Stoll G. "Pitch of vowels: Experimental and theoretical investigation of its dependence on vowel quality", *Speech Communication*, vol.3, pp. 137-150, 1984.

[6] Terhardt, E., Stoll, G. and Seesann M. "Algorithm for extraction of pitch and pitch salience from complex tonal signals" *Journal of the Acoustical Society of America*, vol.71, pp. 679-688, 1982.

[7] Traunmüller H. "Perceptual dimension of openness in vowels", *Journal of the Acoustical Society of America*, vol.69, pp. 1465-1475, 1981.

[8] Katz, W.F. and Assmann, P.F. "Identification of children's and adults' vowels: intrinsic fundamental frequency, fundamental frequency dynamics and presence of voicing", *Journal of Phonetics*, pp. 23-51, 2001.

[9] Mooshammer, C., Hoole P., Alfonso P., & Fuchs S. "Intrinsic pitch in German: A puzzle?" *Journal of the Acoustical Society of America*, 110:2761, 2001.

[10] Pape, D. and Mooshammer, C. "Intrinsic pitch in German - Examining the whole fundamental frequency contour of the vowel" *Proceedings of the CFA/DAGA*, Strasbourg, France.

[11] Honda, K. "Interactions between vowel articulation and F0 control" *Paper presented at LP'98*. Columbus, 1998.

[12] Hoole, P., Honda, K., Murano, E., Fuchs, S. & Pape, D. "Cricothyroid activity in consonant voicing and vowel intrinsic pitch" *Proceedings of the conference on Vocal fold Physiology and Biomechanics*, Marseille, 2004.

[13] Moore, B.C. "An introduction to the psychology of hearing", Elsevier Books, Oxford, UK, 2003.

[14] Soulodre G.A. „Evaluation of objective loudness meters”, Convention paper presented at the 116th AES convention in Berlin, Germany, May8-11, 2004.

[15] Boersma P. and Weenink D. "PRAAT, a system for doing phonetics by computer", *Report of the Institute of Phonetic Sciences of the University of Amsterdam*, pp. 132-182, 1999

[16] Traunmüller, H. "Some aspects of the sound of speech sounds", In: Schouten MEH (ed.) "The Psychophysics of Speech Perception". Dordrecht: Martinus Nijhoff, 293-305, 1987.

[17] Carlson, R., Fant, G. and Granström, B. "Two-formant models, pitch and vowel perception" in: G. Dant and M.A.A.Tatham (Eds.) "Auditory analysis and perception of speech", pp. 55-82, Academic Press:London, 1975.

[18] Kishon-Rabin, L., Amir, O., Vexler, Y. and Zaltz Y. "Pitch discrimination: Are professional musicians better than non-musicians?" *J Basic Clin Physiol Pharmacol*. vol.12, no.2(Suppl.), pp.125-43, 2001.

[19] Patel, A.D., Foxton, J.M. and Griffiths, T.D., "Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech", *Brain and Cognition*, in press

[20] Schon, D., Magne C. and Besson M. "The music of speech: Music training facilitates pitch processing in both music and language" *Psychophysiology*, vol.41, no.3, pp.341-349, 2004.