Is pitch perception and discrimination of vowels language-dependent and influenced by the vowels spectral properties?

Daniel Pape

ZAS Berlin,
Berlin,
Germany.
pape@zas.gwz-berlin.de

ABSTRACT

Pitch discrimination and accuracy has been found to depend on different factors. However, little work has been done (1) on the cross-linguistic influence of the listeners' native language and (2) on the influence of the spectral structure on the pitch perception of vowels as well as (3) cross-linguistic differences regarding different levels of muscical education. If differences in pitch discrimination between different language families exist this would be a crucial knowledge in the design and failure-safe application of auditory displays driven by pitch differences in speech control. Therefore the current study examines pitch discrimination of German vowels with a similar vowel height differing in rounding and tenseness for (1) native German listeners and (2) native Catalan listeners. Significant differences in the sensitivity of pitch perception between these two languages were found. Catalan listeners, independent of their musical education, were mostly insensitive to even large pitch differences in the vowels to be judged. The accuracy of pitch judgements for German listeners were significantly different for musically educated listeners in comparison to musically uneducated listeners. Further, both languages show a significant pitch difference for rounded vowels compared to the unrounded vowels. The current study provides evidence that pitch discrimination is language-dependent, at least partially.

1. INTRODUCTION

The discriminability of complex tones, comparing stimuli differing in pitch, was found to be 2.5Hz [1]. For speech stimuli the discriminability was 1Hz with flat fundamental frequency (F0) contours and 2Hz with rising F0 contours [2]. All values refer to a nominal F0 of 120Hz, the F0 of a male speaker.

It is widely known that accuracy of pitch perception and pitch discrimination of both musical sounds and speech sounds is dependent on several factors. These include:

- musical training, as found in [3] [4] [5],
- spectral structure of the sound, described in [6] ,
- the temporal intensity envelope of the stimuli [7] and
- pitch memory of the listener, as examined in [8].

For speech signals comparing the pitch of different vowels in Germanic languages, it was consistently found that low vowels, i.e. /a:/, have a higher pitch perception than high vowels, i.e. /i:/, when presented with the same fundamental frequency [9][10]. This phenomenon is called "intrinsic pitch" of vowels (IP) and also applies for German diphthongs [11]. However, the reasons and theories for the found pitch difference are widespread:

- Fowler [10], who found a perceived pitch difference of 3.4Hz between high and low vowels, argued that this difference is a compensation to the known "intrinsic F0" phenomenon. It describes consistently a speech production difference in the fundamental frequency of about 10Hz between high and low vowels in nearly all languages [12]. Therefore, following Fowler, the IP is an abstract compensatory mechanism to give a consistent speech melody, independent of the identity of the vowel, which could otherwise disturb "perceptual parsing".
- Stoll [13] claimed pure psychoacoustic reasons for the different perceived pitch between high and low vowels. 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 [14].
- Since the vowel space in Germanic languages is rather crowded (i.e. 15 unreduced vowels for German). Therefore it could be possible that the pitch of vowel is used as a perceptual cue to simplify vowel identification and to avoid perceptual confusion. It was found that the frequency distance between fundamental frequency and first formant is used perceptually to define the "openness" category of a vowel.

What is common in the above cited research is that it only studied pitch perception for listeners of Germanic languages. Since the vowel space of Romance languages is not so crowded compared to the i.e. German vowel space, differences in the pitch discrimination for other language groups could occur. Therefore the phenomenon of IP could be language-dependent, giving different pitch deviations for Romance languages or tone languages.

Surely, for the design of auditory displays it would be crucial to know if pitch discrimination of vowels is language-dependent. This knowledge would be necessary to develop failure-safe and fine-tunable auditory displays which could be applied with the same sound stimuli i.e. in Germany (Germanic language), in Italy (Romance language) and China (tone language).

Therefore the aim of the current study is (1) to compare if pitch discrimination for vowels with a similar vowel height would be different for different language groups. Second, we want to examine the influence of spectral difference on the pitch perception of vowels with a similar vowel height but different spectral slope. Third, we aim at examining the influence of musical education of the listeners on the accuracy of pitch perception and discrimination for the different languages to be studied.

2. METHOD

2.1. The stimuli

The German vowels /i:/, /I/ and /y:/ (comparable to the English vowels in *beat, bid, bur[eau*]) were chosen because they show a similar vowel height and the same intrinsic fundamental frequency [15]. Phonetically, they differ in tenseness, with tense for /i:/ and /y:/ and lax for /I/. Further, due to the roundedness of /y:/ the spectral structure and spectral slope between /i:/ and /y:/ is different, with significant higher energy in higher frequency bands for /i:/ and /I/ due to the different radiation characteristic. 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 context /bVpe/ and were cut from the burst of the preceding plosive to the burst of the following plosive. Since the target word was in stressed position the F0 contours showed a rising slope with a range of 25Hz. The duration was 250ms for /I: y:/ and 125ms for /I/ which is the standard length for these vowels found in literature.

The stimuli were selected in the 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 as well as the virtual pitch values which were computed using the software given by Terhardt [14]. The values indicate that the virtual pitch shift values are identical. Therefore it can be excluded that this psychoacoustic phenomenon could explain any pitch biases which will be measured in the experiment.

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

Table1:Formant values and F2' (perceptual substitute for higher formants, see text section 3.1.) and virtual pitch values (according to the model of Terhardt [14]) for the stimuli of the perception experiment.

Since pitch perception is dependent upon the loudness of the presented stimulus [1] [16], the stimuli were brought to a common loudness. A study of the World Broadcasting Union [17] 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 human perception of loudness and was even superior to complicated psychoacoustic models. Therefore by applying level amplification the stimuli were brought to the same RLB-weighted RMS (SPL_{LEQ}) values (see [17] for en explanation).

Finally, the processed vowel prototypes were pitch-shifted with the PSOLA algorithm with the standard settings of the software PRAAT [18] (see online manual of [18] for references explaining the PSOLA algorithm). The range was set to ±10Hz in 2.5Hz steps (corresponding to a difference of about 2 semitones at 120Hz). The chosen methodology for the pitch 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 his/her judgement. A set with 5 stimuli were presented beforehand to practice the procedure. Three sets with 70 stimuli pairs each were run. Stimuli were paired in randomized order.

2.2. The listeners

The perception experiment was run (1) in Germany with 25 native German listeners and (2) in Tarragona (Catalonia, Spain) with 32 native Catalan listeners. Catalan listeners were chosen to be not educated in the Germanic languages English or German. All listeners were asked for their musical education: 16 of the German listeners and 14 of the Catalan listeners reported musical education. No listeners reported hearing problems.

Additionally to written instructions, listeners were explained the procedure by speaking different vowels in different pitches to make sure they understood the task. Further, 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. “a” either on a piano or on a violin”. Following the practice set, listeners were asked if they had any uncertainties or questions.

3. RESULTS

3.1. Differences in pitch to sound equal in pitch

The response pattern of 8 German and 24 Catalan subjects was markedly nonmonotonic across the continuum and was therefore treated differently (see section 3.3.). For the remaining listeners, we used probit analysis to fit ogives to the curves of individual subjects. Our dependent measure 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 opportunities.

Table 2 gives the mean and significance values for the German and Catalan listeners, split by musical education. Only for /i:/-/y:/ a significant departure from 0Hz for the German musically educated listeners was found. For the Catalan listeners, the pitch comparison of the vowel pair /i: I:/ reached significance for the Catalan musically educated listeners. The values for the Catalan musically uneducated listeners were not analyzed due to their small sample size.

	Mean /i:I/	Mean /i:y:/	Significance /i:/I	Significance /i:y/
German musically educated	0.89	-3.92	P=0.571	**p= 0.02**
German musically uneducated	1.91	-5.26	P=0.513	P=.356
Catalan musically educated	3.76	-9.23	**P=.048**	**P=.029**

Table2: Mean values and significances of a t-test for German and Catalan listeners, split by musical education. Significant values are printed in bold.

Thus, as was also found by Stoll [13], spectral shape, i.e. lip rounding, has a significant effect on pitch judgements, with significant lower pitch responses for the rounded vowel compared to the unrounded vowel. This effect was significant in both language families, with even lower pitch responses in the Romance language. Opposite to the theory of Stoll and Terhardt, in our data, "virtual pitch shift" could not explain this significant difference in pitch judgements.

This influence of the spectra on the pitch judgement could be explained with an interaction of "sibilant pitch" which is the perceived pitch of whispered vowels: Traunmüller [19] found that this pitch (which occurs therefore in absence of glottal vibrations) corresponds to F2', an average of the higher formants starting from F2. In presence of glottal vibrations, Carlson [20] found with two-formant models that the F2' for Swedish /i:/ is 3210Hz, compared to 2010Hz for /y:/ (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 control if listeners are biased to judge more fundamental pitch or more sibilant pitch, an interference of sibilant pitch and fundamental pitch 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.

3.2. Dependence of the of pitch responses on musical education

Figure 1 gives the means of the responses "/i:/ is heard higher in pitch" in dependence of the presented F0 difference, split by musical education. A pitch experiment of simple complex musical tones differing only in pitch is indicated by a linear rising function of about 45 degrees in figure 1.

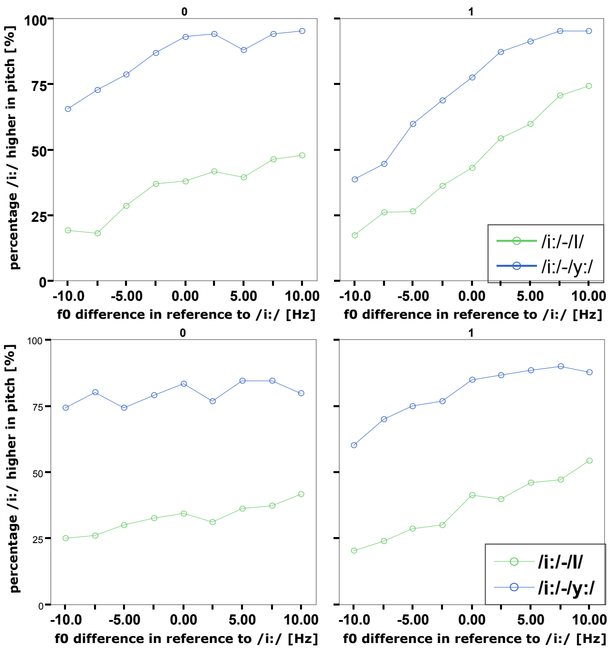

Figure 1: Mean values for the judgement "/i:/ is higher in pitch" in dependence of the F0 difference of the stimuli, split by musical education (0=musical uneducated). In dark colour the differences /i:/-I:/ and in light colour the differences /i:/-/y:/ are shown. The upper panel shows the German results, whereas in the lower panel the results for the Catalan listeners are given.

As can be seen in the figures, all response functions are parallel for both vowels to be judged. The German musically educated listeners follow more precisely the given F0 difference compared to the musically uneducated listeners, which is in accordance with results found in literature [3] [4] [5]. For Catalan, this response difference for the given F0 difference can hardly be found. The results for the musically educated Catalan listeners tend to show the same or lower linear rising functions as the musically uneducated listeners for German. For both musically educated and musically uneducated listeners, Catalan listeners tend to judge the pitch of the vowels more categorically, i.e. they judge the vowel /I/ higher in pitch compared to the vowel /i:/, independent of the given F0 differences.

3.3. General Sensitivity to pitch differences: German vs. Catalan

To examine pitch judgements with regard to the F0 difference, the regression lines between the judgements of each listener that /i:/ was heard higher in pitch (collapsed over both vowels) and the corresponding F0 difference was computed. The regression lines are shown in figure 2. A steep regression line would indicate a consistent response to the given F0 difference, whereas a more horizontal line would indicate a low sensitivity to F0 difference. In this case the response of the listener is more influenced by the identity of the vowel to be judged, independent of the increase of the physical F0 difference.

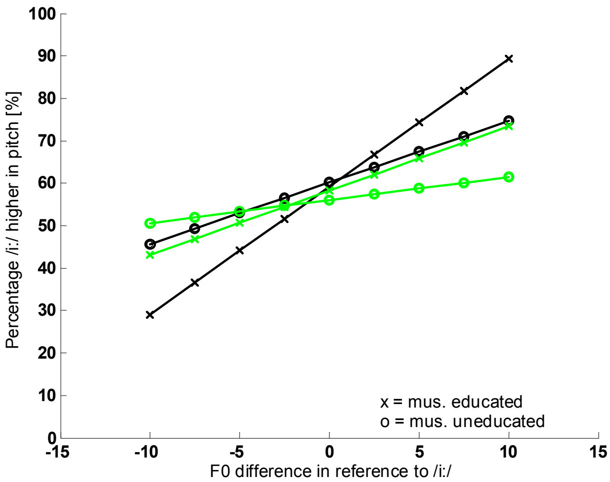

Figure 2: Regression lines collapsed over all listeners and split by musical education. In dark colour the line for German listeners, in light colour the lines for Catalan speakers are shown.

As can be seen in figure 2, the regression line for the German musically educated listeners is the steepest. Surprisingly, the line for the Catalan musically educated listeners is not as steep as the one corresponding to German musically uneducated listeners, indicating insensitivity to the pitch difference to be judged.

4. DISCUSSION AND OUTLOOK

The present study examined possible pitch differences for German vowels with a similar tongue height. The study was conducted with listeners of a Germanic language and listeners of a Romance language to examine possible cross-linguistic differences. Any found differences could not be due to psychoacoustic pitch shift which in the literature is found as one of the possible reasons for intrinsic pitch differences [13].

Both languages show significant pitch differences for the pair differing in roundedness (and therefore in higher frequency radiation), which can be explained by an influence of the "fundamental pitch" judgements with a "sibilant pitch" judgement due to the difference in the higher formants between these two vowels.

Further, results indicate a systematic difference in pitch perception and discrimination for Germanic languages compared to Romance languages This difference is not due to musical education of the listeners. It should be expected that musical education improves the sensitivity to pitch differences. However, with most musically educated Catalan listeners an insensitivity to pitch differences of 2 semitones was observed. Regression lines between F0 difference and pitch judgements for musically educated Catalan listeners indicate the same pitch difference sensitivity as for German musically uneducated listeners.

It is not clear what mechanisms cause this reduced pitch discrimination insensitivity of the Catalan musically educated listeners for the given F0 differences to be judged. The results could indicate a suppression of the musical pitch processors and a take-over of the prosodic pitch processors, otherwise the Catalan musically educated listeners should show the same sensitivity to the given F0 differences as the German musically educated listeners. In neuroscience it is still not clear if separate pitch perception mechanisms for speech prosody and musical melodies exist. On the one hand researchers, i.e. Patel [21], give neuroimaging evidence for different processing for speech prosody and musical melody. On the other hand researchers, i.e. Besson [4], provide 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 author's knowledge such a comparison has not been published. 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. I 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 I am thankful to B. C.Moore, H.Traunmüller, W.Serniclaes and J.Brunner for helpful comments.

6. REFERENCES

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

[2] D.H.Klatt, "Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception", *Journal of the Acoustical Society of America*, no.53, pp.8-16, 1973.

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

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

[5] F.H.Rauscher and S.C.Hinton, "Type of music training selectively influences perceptual processing", *Proceedings of the European Society for the Cognitive Sciences of Music*, Hannover, Germany, 2003.

[6] G.Stoll, "Spectral-pitch pattern: A concept representing the tonal features of sounds", in M. Clynes (ed) "Music, Mind, and Brain", pp.271-278, Plenum Press, New York, 1982.

[7] A.J.M.Houtsma and T.D.Rossing, "Effects of signal envelope on the pitch of short complex tones" *Journal of the Acoustical Society of America* , vol.81, no.2, pp.439-444, 1986.

[8] N.Gaab and G.Schlaug, "The effect of musicianship on pitch memory in performance matched groups" *NeuroReport,* vol.14, no.18, pp.2291-2295, 2003.

[9] K.Chuang and W.S.Y.Wang, "Pscychophyiscal 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.

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

[11] O. Niebuhr, "Intrinsic Pitch in Opening and Closing Diphtongs of German", in B.Bel, I. Marlien (Eds.) *Proc. International Conference: Speech Prosody 2004*, Nara, Japan, 2004.

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

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

[14] E.Terhardt, G.Stoll, and M.Seesann, "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.

[15] C.Mooshammer, P.Hoole, P.Alfonso and S.Fuchs "Intrinsic Pitch in German: A Puzzle*?", 142nd meeting of the Acoustical Society of America* in Ft. Lauderdale, Florida, 3-7 december, 2001.

[16] E.Zwicker and H.Fastl, *Psychoacoustics: Facts and Models*. Springer-Verlag, Heidelberg, Germany, 1990.

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

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

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

[20] R.Carlson, G.Fant and B.Granström. "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.

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