

Language background affects the strength of the pitch bias in a duration discrimination task

Daniel Aalto¹, Juraj Šimko², Martti Vainio¹

¹University of Helsinki, Finland

²Bielefeld University, Germany

daniel.aalto@helsinki.fi, juraj.simko@uni-bielefeld.de, martti.vainio@helsinki.fi

Abstract

The fundamental frequency of a complex sound modulates the perceived duration of a sound. Higher pitch sounds are perceived longer compared to lower pitch sounds as shown by several independent studies since 1973. In this paper, the effect of language background is studied: native speakers of Finnish and German participated in a two alternative forced choice duration discrimination experiment where the duration and frequency of two sounds are randomly varied. The overall duration discrimination sensitivity was similar to both groups but the speakers of Finnish were influenced more by the pitch in their judgements. In addition, the difference in the two sounds' pitch period explained the response data better than the difference in pitch frequencies or the pitch interval. As the Finnish quantity system is known to employ both duration and pitch cues, the present results suggest that the speakers are shaped by the language environment even when the task is purely non-linguistic.

Index Terms: duration discrimination, pitch, Finnish, German

1. Introduction

When listeners are asked to compare duration of two sounds with static f_0 contour, they tend to judge the one with a higher pitch as longer. This phenomenon has been confirmed by multiple investigation involving comparisons of sounds of different types (from pure tones to speech segments) and of a great range of fundamental frequencies (from 100 Hz to 8 kHz) [1, 2, 3, 4, 5].

In the beginning of 1970's, Burghardt asked German subjects to adjust a duration of a sinusoidal with different pitch than the standard. Based on the comparisons between several sinusoidals varying between 200 Hz to 8 kHz, Burghardt showed that 3 kHz sounds were perceived the shortest. He concluded that the frequency might on the one hand influence the perceived duration by an additive constant proportional to the pitch period. On the other hand, the processing speed might vary over the different spectral bands of the basilar membrane filters. Moreover, Burghardt observed a similarity between equal duration curves and equal loudness curves [1]. The latter conclusion fits well to Frankenhaeuser's theory of duration perception since louder sounds are perceived longer, the reaction times for the sound onset detection decrease with increasing intensity, and the temporal discrimination is more precise for louder sounds [6, 7, 8, 9, 10, 11]. The former conclusion is supported by observations from neurophysiology: first, the entrainment to the pitch of trapezoid body cells relates the pitch rate and total activity of the neural population to the pitch frequency [12]; second, the response of the cells in the central nucleus of inferior colliculus that are sensitive to the amplitude modulated

sinusoidals, has a latency proportional to the pitch period [13]

In 1977, Rosen confirmed that the pitch affects the duration judgements of complex sounds as well (although, apparently, he hadn't been aware of the work of Burghardt) but offers no explanation for the observations. Seemingly unaware of all the earlier work, Brigner offered a new explanation in his experiment on the apparent duration difference of sinusoids at 0.5 kHz and 4 kHz: the higher pitch sound is perceived as smaller and hence as longer [3]. However, the Brigner's results could be explained with varying loudness: a 4 kHz sound is louder than a 0.5 kHz sound and hence perceived longer.

A closely related question is the difference limen (DL) for the sound duration i.e. what is the smallest ΔT such that a stimulus with duration T can be distinguished from another stimulus with duration $T + \Delta T$ other stimulus features being equal. According to Abel [9], the DL is roughly a tenth of the stimulus duration for stimuli with durations above 50 ms. The ratio between difference limen and the stimulus duration is called Weber fraction. The Weber fractions decrease as a function of stimulus duration and these fractions are smaller for musicians. In addition, when the stimuli have varying pitch, the duration discrimination deteriorates. Jeon and Fricke propose that the reason for the phenomenon lies in the different integration times involved in the frequency processing: the higher pitch sound would have faster detection time and hence a longer duration [4]. However, the loudness differences could have confounded the results.

The magnitude of the effect that pitch has on perceived duration has been estimated to be unimportant for speech [2, 14]. However, more complex stimuli with several pauses and bursts have shown that the perceived duration difference of a longer sequence can be strongly modulated by the pitch [15, 16]. Moreover, in tonal languages like Thai and Cantonese, the high tones are produced in average shorter than low tones [17, 18]. This can be regarded as (over) compensation for the temporal asymmetry in the perception of the durations [5]. If the small duration distortion caused by pitch would be indeed negligible with respect to speech, then the language background of a speaker would not be expected to affect the strength of the effect. However, the native language is known to affect the early neural processing of pitch and duration [19, 20, 21].

In this paper, we revisit the pitch-duration interaction and investigate it from two perspectives: first, we attempt to falsify the pitch period mechanism proposed by Burghardt by examining the key predictions of the proposed account: the degree to which the pitch influences perception of duration would be *linearly* proportional to the difference in *period* of the two compared sounds if the sounds have equal loudness. As a consequence, it would be non-linearly proportional to pitch frequency and pitch interval. Second, the relevance for speech is assessed

by analyzing if language affects the phenomenon by comparing the responses of the speakers of two languages, Finnish and German. In Finnish, a language with full-fledged quantity system, the phonological length of the vowel is signaled by the duration and the pitch while in German, a more restricted quantity system takes place where the length opposition is signaled by vowel quality and duration [22, 23, 24, 25, 26, 27].

2. Methods

A two alternative forced choice duration discrimination task with 300 static pitch pairs and 300 dynamic-static pairs were presented for native speakers of Finnish and German. Here we only describe and report the results for the static pairs. There were 12 Finnish (11 females; age 25–67 years, mean 37) and 6 German (3 females; age 24–41 years, mean 31) participants. The participants reported no hearing problems. Informed consent of all subjects was obtained and the experimental arrangement was in accordance with the declaration of Helsinki.

2.1. Stimuli

The stimuli were varied in three dimensions: the actual duration, the pitch level, and the pitch movement. Half of the stimuli were level-level pairs which are described now in detail.

The first sound in the pair is called a and the second x . Each stimulus pair (a, x) consisted of a short and a long sound: the short had a random duration between 100 and 150 ms; the long was randomly 0–50 ms longer than the short one. The long and short durations were attached randomly to a and x .

One of the sounds in the pair (a, x) was always low and the other one high. The low tone had a random period of 7.5–15 ms (corresponding to ca. 67 and 133 Hz). The higher tone had randomly 0–5 ms shorter period (frequency range of 67 and 400 Hz). The high and low pitch values were attached randomly to a and x independent of their durations.

The onsets of the sounds had a fixed difference of 800 ms. The sound signals were constructed based on simple sawtooth waves. These waves were then gamma filtered with center frequency 3000 Hz. Finally, the energy of the signal was normalized so that the intensity of the sound signals integrated over the first 100 ms was equal. This lead to approximately equal loudness.

2.2. Procedure

The experiment was run on Matlab on a laptop. The participants heard pairs of sounds through head-phones in a sound proof studio and had to choose if the first or the second sound in the pair was longer by typing “a” respectively “x”. A few practice sounds with duration ratio 2 or 0.5 and equal pitch static sounds were presented to familiarize the participants with the task and to let them adjust the amplitude of the head-phone signal. The typing of the answer prompted the generation of the next stimulus and it was played to the participant after a delay of 500 ms. The participants were told to concentrate on the duration and neglect any other variation in the stimuli.

2.3. Analysis

A mixed effects logistic regression model was fit to the data to estimate the psychometric function for duration discrimination in the responses. The fixed effects were actual duration difference d , the period difference of the pitches p , and the mother tongue of the speaker $L1$ together with the first order interac-

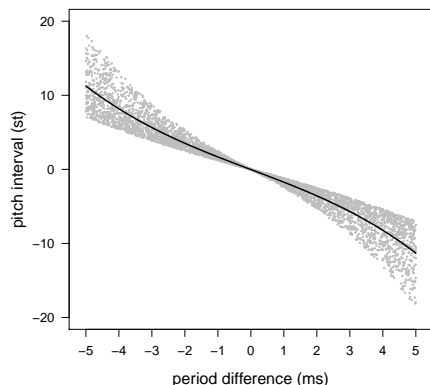


Figure 1: Structure of the experimental design is revealed by plotting the pitch interval between the sounds as a function of their pitch period difference. The solid line depicts the best third order polynomial regression model between these variables.

tions. The subjects were treated in the model as random effects with random slopes for duration differences. Note that every variable included in the analysis is antisymmetric with respect to the order, i.e. if the order of the two sounds is reversed, then the value of the variable changes the sign. A selection between models was done using anova comparisons by reducing the full model by removing the least significant terms. This led to the minimal model called *linear period difference model*.

The same process of data fitting was then done again with period difference substituted by pitch interval in semitones. Again the anova was used for finding the minimal model, *linear pitch interval model*. The *linear frequency difference model* was constructed similarly. The three different models were compared using the proportional reduction in deviance against the null model which included only random intercepts for the subjects. Moreover, the linearity of the models was assessed by introducing quadratic and cubic terms and analyzing their impact to the deviance reduction.

Since the stimuli were generated with several random parameters, the obtained data set offers a possibility to compare the effect of other pitch related variables. In Figure 1, the dependence between the pitch period difference and the pitch interval is demonstrated. A third order polynomial regression line was fitted to connect the two variables. The regression model explained 93.5% of the total variation. Hence, the third order polynomial essentially captures the dependency between the two variables in the experimental design.

To compare the results to earlier results on difference limens, the DL was defined by estimating the 75% correct discrimination threshold from the logistic regression model. The estimate was calculated by deleting the terms not related to duration leading to $DL = \ln(3)/k$, where k is the slope of the duration term in the logistic model.

3. Results

The duration difference was the most important single factor in every model analyzed. The model with duration difference as the only explanatory factor reduced the deviance by 23.5 %. In the following we describe the influence of the pitch to the

responses.

3.1. Comparison of models

First, we compare the resulting minimal models using three pitch-related difference measures as fixed effects. The measures used were *period difference*, *frequency difference* and *pitch interval*, respectively. All three mixed effect logistic regression models showed similar significance patterns regarding main effects and interactions. In all three cases, the main effects of actual *duration difference* and of the pitch-related measure were both significant, as was the intercept. The interaction between the pitch-related measure and *L1* was also significant for all three models. In the model using *period difference*, the interaction between this measure and *duration difference* was also significant, but this was not the case for the other two frequency-related measures of pitch difference. No other effects were significant in either of the models, including the main effect of *L1*. Only significant main effects and interactions were used in the minimal models.

The minimal models with period difference, frequency difference and pitch interval reduced the variance by 30.1 %, 27.4 % and 29.0 %, respectively, relative to the null model. The interaction that was only significant for the period difference measure and therefore included in the minimal model (*duration difference* \times *period difference*) was not fully responsible for the difference; the pitch difference model without this interaction still reduced the null variance by 30.0 % (nevertheless, as shown by ANOVA-comparison between the models, this tiny decrease is significant).

The mixed effect logistic regression model using *period difference* as a measure of pitch difference between stimuli provided the greatest explanatory power. It is important to stress that the small differences between variance reduction percentages are expected, as the three measures compared are highly correlated.

3.2. Linearity of the effects

Both quadratic (symmetric) and cubic (antisymmetric) terms were introduced to analyze the nonlinearity effect in the *duration difference* and the measures of pitch difference: *period difference*, *frequency difference*, and *pitch interval*, respectively. Since the quadratic terms had in all respects negligible impact on the responses, only cubic terms are reported here.

The actual *duration difference* had linear effect on the responses: The cubic term was not significant ($z < 2$), the model was not significantly better than the linear model (ANOVA), and adding the cubic term decreased the deviance by 0.01 % only.

The *period difference* effect was also mostly linear. While the cubic term was significant $z = 2.0$ and the linear model was significantly different from the cubic model (ANOVA, $p = .02$), the additional reduction in deviance introduced by non-linearity was only 0.08 % (the explanatory power was not improved and stayed 30.1 % as for the linear model).

The *frequency difference* model is most substantially improved by adding the nonlinear term: the cubic term was significant ($z = -11$), the non-linear model was significantly different from the linear model, and the additional reduction in deviance was 2.2 % (total reduction was 29.6 %).

The *pitch interval* model improved similar to frequency difference model by adding the nonlinear term: the cubic term was significant ($z = -10$), the non-linear model was significantly different from the linear model, and the additional reduction in deviance was 1.1%. The resulting non-linear model reduced

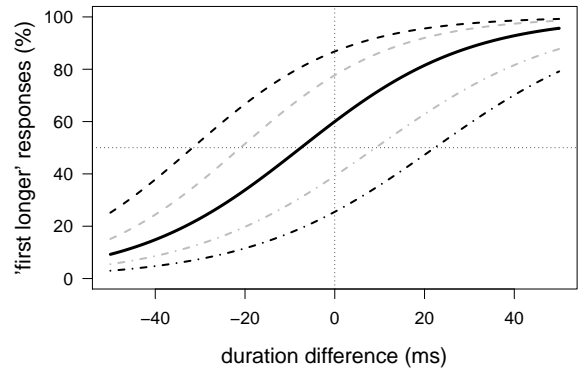


Figure 2: The logistic curves fitted to the response data. The proportion of the responses (ordinate) is plotted as a function of duration difference (abscissa) for different period difference levels: solid line (same pitch), dashed line (first sound higher), and dot-dashed line (first sound lower). The grey curves represent the German and the black curves the Finnish participants' responses.

the variance by 30.5% compared to the null model and thus provides an explanatory power similar to the linear *period difference* model.

The similarity between the linear *period difference* and non-linear *pitch interval* models is, of course, not surprising. These two measures are highly correlated and related to each other in a non-linear fashion, see Figure 1. The results reported so far suggest that the influence of the pitch difference on durational judgements of participants is approximately linearly proportional to the difference in period, and (perhaps consequently) non-linearly to derived frequency measures. As the revealed interaction patterns suggest, the strength of the influence is language dependent. To analyze this language dependence, we shall use the simpler, non-linear *period difference* model.

3.3. Language specificity of the influence of period difference

The minimal mixed effects logistic regression model using *period difference* is summarized in Table 1.

Table 1: The minimal mixed effects logistic regression model with linear *period difference* term.

Fixed effect	Estimate	z value
Intercept	0.406	4.1
Duration difference	53.8	14
Period difference	-296	-18
Duration difference:Period difference	-1130	-2.3
Period difference:L1 (German)	127	4.8

To help interpret the model parameters, several resulting logistic regression curves are plotted in Figure 2. The percentage of responses (“first longer”) is shown as a function of actual duration difference between the sounds (positive difference means that the first sound was longer).

The solid line is an estimate of durational judgements for

both Finnish and German speakers for stimuli with equal pitch (in this case the responses were statistically indistinguishable due to the lack of significant main effect of *L1* and its interaction with *duration difference*). In this case, when the sounds were of equal duration, the first sound was perceived as longer (significant intercept). To achieve balanced judgement, the second sound should be approximately 9 ms longer than the first one.

The significant interaction between *L1* and *period difference* signals the different magnitude of pitch effect on durational judgement between the speakers of the two languages. The dashed lines (black for the Finns and grey for the Germans) show estimates for the situation, when the period of the first sound was 5 ms *shorter* than that of the second sound, i.e., the first sound was of a higher pitch. As we can see, in this case the first sound was considerably more likely to be perceived as longer by Finnish participants than by their German counterparts. In order to compensate for such a great difference in pitch (the maximum used in our stimuli), the second, lower pitch sound must have been approximately 35 ms longer than the first sound for the Finns, but only about 20 ms longer for the Germans. Similar phenomenon occurs when the pitch of the second stimulus was higher: to achieve balanced responses, the lower-pitch sound should be approximately 25 ms longer for Finnish participants but less than 10 ms for German ones.

3.4. Duration difference limens

To analyze the duration difference limens, the pitch period based logistic regression models are used. The estimate for the smallest perceivable difference is 20 ms. The mean duration of the stimuli was 137 ms and hence an estimate for the Weber fraction is 15 %.

4. Discussion

The aim of this study was to investigate in more detail the earlier reported pitch-duration interaction. The responses were analyzed for the linearity/non-linearity of the explanatory variables and for native language effect. The results show that the pitch effect is best described in terms of the pitch period difference of the two sounds. This interpretation is however not without reservations: the experimental design resulted in low frequencies to be over represented in the trials. In 67% of the sound pairs, both sounds had a pitch in the one octave range from 67 Hz to 133 Hz. Nevertheless, the highest pitch sounds extended to 370 Hz. Despite the imbalance, the pitch distribution is roughly comparable to the distribution of male fundamental frequencies in spontaneous speech.

As reflected by the reductions in deviance, the duration judgements were based mostly on the actual differences in the sound durations. Earlier, Finnish has been found to be more precise in their duration discrimination [20], but this effect was not statistically significant in the current data. However, there is no contradiction since the reported values for the effect size are much below the response variance here.

The studies of this phenomenon, including ours, report *judgements* of participants, not perception *per se*. Moreover, the auditory processing steps are not available by a behavioral study. Hence, any implications to neural processes remain suggestive and speculative. However, the results point to a prominent role of pitch period in explaining the phenomenon as opposed to other pitch related variables as pitch frequency and pitch interval, at least in the low frequency range.

As opposed to earlier studies, the loudness level was more carefully controlled for by restricting the energy of the sound signal to a narrow frequency band (to minimize the sensitivity differences across spectrum) and normalizing the intensity. However, the varying durations and pitch levels might have created variations in the loudness sensations.

Finally, the results support the Burghardt's view that the pitch modulates the perceived duration proportional to the duration of pitch period. In addition, the effect is not due to basilar membrane mechanics (ringing) since the stimuli had equal spectral distribution and hence the basilar membrane responses are expected to be identically located. Also, the pitch bias has relatively large impact on duration discrimination (see Section 3.3) varying among individuals in a language specific way.

5. Conclusions

The interaction of duration and pitch was studied using a two alternative forced choice duration discrimination task. As previously found, the higher pitch was systematically associated with longer duration judgement. A detailed analysis of the response data showed that the phenomenon depends rather linearly on the pitch period difference of the two sounds. This suggests that the cause of the phenomenon could be related to the peripheral auditory processes: the pitch processing contributes to the total amount of activation proportional to the frequency hence shortening the onset detection latency for higher tones.

All the participants were native speakers of Finnish or German. Surprisingly, the pitch had a much stronger effect on the duration judgements of the Finnish than those of the Germans while the overall duration discrimination accuracy was comparable between the groups. Given that the quantity is signaled by duration and pitch cues in Finnish, the pitch processing pathways in the auditory system might have been more activated during the duration discrimination task in the Finnish participants explaining the between group difference. To rule out this explanation, another study is needed which directly measures the early auditory processes.

Finally, there seems to be a universal tendency for higher pitch sounds to be perceived longer when all the other sound parameters are kept constant. During language acquisition, the native language shapes the developing brain, and the ways in which phonological distinctions are signaled become hard-wired in the population. The implication for understanding the language change is fundamental: the language shapes the perceptual system of the new generation of speakers modifying the phonetic landscape where the phonological distinctions are realized. Hence, the adaptation of the perceptual system to the language environment might be partly driving the language change.

6. Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 287678. This work was partially supported by the Alexander von Humboldt Fellowship grant to the second author.

7. References

- [1] Burghardt, H., "Die subjektive Dauer schmalbandiger Schalle bei verschiedenen Frequenzlagen", *Acustica*, 28:278–284, 1973.
- [2] Rosen, S. M., "The effect of fundamental frequency patterns on perceived duration", *STL-QPSR*, 18(1):17–30, 1977.
- [3] Brigner, W. L., "Perceived duration as a function of pitch", *Perceptual and Motor Skills*: 67:301–302, 1988.
- [4] Jeon, J. Y. and Fricke, F. R., "Duration of perceived and performed sounds", *Psychology of Music*, 25:70–83, 1997.
- [5] Yu, A. C. L., "Tonal effects on perceived vowel duration", in C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée [Eds.], *Laboratory Phonology 10*. Berlin, Mouton de Gruyter 2010.
- [6] Frankenhaeuser, M., "Estimation of time", *Almqvist & Weksell*, Stockholm, 1959.
- [7] Berglund, B., Berglund, U., Ekman, G. and Frankenhaeuser, M., "The influence of auditory stimulus intensity on apparent duration", *Scand. J. Psychol.*, 15(1):277–284, 1974.
- [8] Aurell, C. G., "Temporal integration and perception of duration", *Scand. J. Psychology*, 15(1):277–284, 1974.
- [9] Abel, S. M., "Duration discrimination of noise and tone bursts", *J. Ac. Soc. Am.*, 51:1219–1223, 1972.
- [10] Stebbins, W. C., "Auditory reaction time and the derivation of equal loudness contours for the monkey", *J. Experimental analysis of behavior*, 9(2):135–142, 1966.
- [11] Henry, F., "Discrimination of the duration of a sound", *J. Exp. Psychol.*, 67: 483–485.
- [12] Joris, P. X., Louage, D. H., Cardoen, L. and van der Heijden, M., "Correlation index: a new metric to quantify temporal coding", *Hear. Res.*, 216: 19–30, 2006.
- [13] Joris, P. X., Schreiner, C. E. and Rees, A., "Neural processing of amplitude-modulated sounds", *Physiol. Rev.*, 84:541–577, 2004.
- [14] Lehiste, I., "Influence of fundamental frequency pattern on the perception of duration", *J. Phonetics*, 4:113–117, 1976.
- [15] Burghardt, H., "Über die subjektive Dauer von Schallimpulsen und Schallpausen", *Acustica*, 28:284–290, 1973.
- [16] Zwicker, E. and Fastl, H., "Psychoacoustics: facts and models", Springer-Verlag, Heidelberg, 1990.
- [17] Gandour, J., "On the interaction between tone and vowel length: Evidence from Thai dialects", *Phonetica*, 34:54–65, 1977.
- [18] Kong, Q., "Influence of tones upon vowel duration in Cantonese", *Language and Speech*, 30(4):387–399, 1987.
- [19] Krishnan, A., Xu, Y., Gandour, J. and Cariani, P., "Encoding of pitch in the human brainstem is sensitive to language experience", *Cogn. Brain Res.*, 25:161–168, 2005.
- [20] Tervaniemi, M., Jacobsen, T., Röttger, S., Kujala, T., Widmann, A., Vainio, M., Näätänen, R. and Schröger, E., "Selective tuning of cortical sound-feature processing by language experience", *Europ. J. Neuroscience*, 23:2538–2541, 2006.
- [21] Marie, C., Kujala, T. and Besson, M., "Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds", *Cortex*, 48(4):447–457, 2012.
- [22] Suomi, K., "Temporal conspiracies for a tonal end: segmental durations and accentual f₀ movement in a quantity language", *J. Phonetics*, 33:291–309, 2005.
- [23] Vainio, M., Järvikivi, J., Aalto, D. and Suni, A., "Phonetic tone signals phonological quantity and word structure", *J. Acoust. Soc. Am.*, 128:1313–1321, 2010.
- [24] Järvikivi, J., Vainio, M. and Aalto, D., "Real-time correlates of phonological quantity reveal unity of tonal and non-tonal languages", *PLoS one*, 5(9):619–639, 2010.
- [25] Wiese, R., "The phonology of German", Clarendon Press, Oxford, 1996.
- [26] Strange, W. and Bohn, O.-S., "Dynamic specification of coarticulated German vowels: perceptual and acoustical studies", *J. Acoust. Soc. Am.*, 104(1):488–504, 1998.
- [27] Kohler, K. J., "Stressed vowel duration in German: Results from a large data base of read speech", *Acta Ling. Hafniensia: International J. of Linguistics*, 27(1):299–321, 1994.