Intrinsic F0 differences for German tense and lax vowels

Daniel Pape[1*], Christine Mooshammer[2]

[1]ZAS – Centre for General Linguistics
Jägerstrasse 10-11, 10117 -Berlin-Germany

[2]IPDS, Institut für Phonetik und Digitale Sprachverarbeitung
CAU, 24098 Kiel-Germany

pape@zas.gwz-berlin.de, timo@ipds.uni-kiel.de

Abstract. *This study examines the vowel-intrinsic fundamental frequency (IF0) for tense and lax vowels in German: These show a similar F0 but differ in their articulatory vowel height. This is contrary to commonly accepted biomechanical theories explaining IF0 by a physiological linkage between the upper vocal tract and the larynx. To examine this phenomenon, acoustic, electroglottographic and articulatory data were recorded for three German speakers producing the tense vowels /iː uː aː/ and their lax counterparts. Results for vowel mid, onset and offset showed that the articulatory positions significantly differed for tense-lax vowel pairs but the measured F0 was similar. A stepwise regression procedure selected the vertical tongue back sensor to explain the largest amount of the variance between the vowel F0 values. Correlation and regression values of both tense and lax high versus low vowels were high and significant. In contrast, regarding the tenseness difference, correlation and regression analyses between the vertical tongue back sensors and the corresponding F0 values were weak and non-significant. Thus, biomechanical IF0 theories can not account for the described tenseness phenomenon. Other mechanisms seem to be responsible, i.e. an active rise of F0 by the speaker to differentiate between tense and lax vowels.*

1. Introduction

The present work studies the relationship between fundamental frequency data (F0) and articulatory data (EMMA). More specifically, the phenomenon "Intrinsic fundamental frequency" (hereafter IF0) will be investigated with special reference to German tense and lax vowels. Generally, IF0 describes the relationship between F0 and phonological vowel height: Close vowels show a higher F0 compared to open vowels other things being equal. The possible reasons causing IF0 differences are controversially discussed: On the one hand researchers argue that IF0 is an automatic consequence of the connection between the tongue and the laryngeal structures (see i.e. Ohala and Eukel

[*]Supported by a grant from the German Research Council GWZ 4/8-1 P1.

1987), thus the speaker automatically increases F0 when articulating close vowels due to the physiological linkage between the articulatory structures (tongue pull). In contrast, enhancement theories (i.e. Diehl and Kluender 1989) claim an increase in the robustness of the vowel identification: IF0 differences may serve as a vowel quality cue to robustly enhance the discriminability between different vowels or vowel classes (based on Traunmüller 1984). In the last decades experimental evidence has been presented in support of both theories. It shall be noted that most researchers claim the automatic consequence theory to be more reliable, mostly due to the strong evidence given by work of Whalen and Levitt (1995): They found IF0 to be present in all major language families. This result thus strongly challenges the enhancement theory, because languages with a three or five vowel contrast do not need an enhancement of the contrast, since vowels are already widely separated in a perceptual context.

However, for the tense and lax vowels in German it was found that these vowel classes show a similar acoustic F0 although their articulatory vowel height significantly differs (Fischer-Jørgensen, 1990; Mooshammer et al. 2001). Thus, this finding partly contradicts the automatic consequence theory since according to the biomechanical link these vowels should show a differing F0. It could be argued that the articulatory differences are too small to induce IF0 differences. However, the findings in the literature speak against this assumption: IF0 differences are found to gradually increase with increasing vowel height and are significantly different even for small articulatory distances, i.e. /iː/ versus /eː/ in German (Fischer-Jørgensen, 1990). However, the reason for the unexpectedly high IF0 in lax vowels could be explained by EMG data from Hoole et al. (2004): Higher cricothyroid (CT) activity was found for German lax vowels compared to tense vowels. CT is one of the main muscles influencing F0, therefore it could be concluded that the speakers actively influence the F0 of lax vowels for a certain reason. This would explain why F0 is not lower as would be expected according to the tongue pull theory.

The aim of the present study is to extent the previously described results in the literature with the examination of combined F0 and articulatory recordings. Using this setup, a link between these dimensions as proposed by the automatic consequence theory should reveal a strong relationship when comparing the corresponding data. One shortcoming of previous studies was that only the midpoints of the vowels were examined. However, if IF0 is an automatic consequence, then F0 contours during the whole vowel and movements of one or more articulators should be correlated with each other. Thus, as a simplification additionally also the vowel onset and the vowel offset will be examined.

2. Methods

Combined acoustic and physiological data were recorded by means of synchronized acoustics, EGG and EMMA (Electromagnetic Midsagittal Articulography, Carstens Medizinelektronik AG100). We regarded the time intervals between the sharp peaks in the derivative of the EGG signal (corresponding to the rapid glottal closure) as the temporal periods on which F0 calculations were based. These correspond better to glottal excitation and are shown to be more robust than different tested F0 extraction methods based on the audio signal. All peaks were manually examined to ensure that no double peaks would disturb or decrease reliability of the F0 measures.

Tongue, jaw and lower lip movements of three male speakers of Standard German were recorded by EMMA. Three sensors were attached to the tongue, one as far back as possible (tongue back: BK), one approximately 1 cm behind the tongue tip (tongue tip: TP) and one equidistant between BK and TP (tongue blade: BL). One sensor at the lower incisors was used for tracking jaw movements. The EMMA signals were downsampled to 200 Hz and low-pass filtered at 30Hz. The recorded corpus consisted of the German tense and lax vowels /iː ɪ uː ʊ ɑː a/ in nasal context to avoid distortion of the fundamental frequency contour due to devoicing. The carrier sentence was “Sage mVme bitte.” (Say mVme please) and all sentences were repeated 8 times in randomized order. Measured acoustical landmarks included the onset of the target vowel, the mid of the vowel and the offset of the target vowel. Vowel mid was defined as the temporal midpoint between the corresponding acoustic vowel onset and offset.

For statistical analyses we used bivariate correlation (Pearson) with a two-tailed test for significance and linear regression (both inclusion and stepwise procedures) from SPSS version 11.5. For labelling and processing of the synchronized EGG and acoustic signal PRAAT and Matlab was used.

3. Results and Discussion

3.1. Fundamental frequency

In order to test whether our data confirm results from earlier studies we first measured F0 values at the vowel mid. Figure 1 displays boxplots separately for each speaker for all examined vowels.

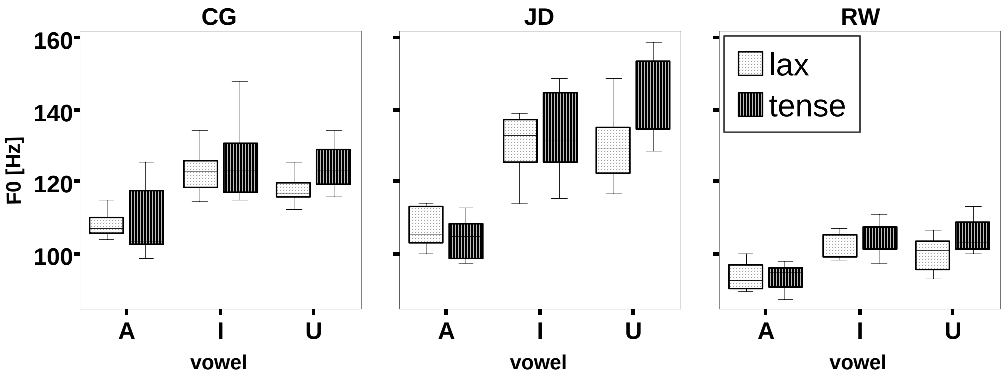

Figure1: Boxplots of the F0 values at the vowel mid for all examined vowels, given separately for each speaker. Black boxes indicate tense vowels, white boxes indicate lax vowels.

As can be seen, the high vowels are about 15-20Hz higher in comparison to the low vowels for all speakers. For speaker RW generally all F0 values are lower with a difference of about 10Hz, but showing the same trend as the other speakers. Further, it can be seen that when comparing the F0 of each tense and lax vowel pairs, the F0 values are rather similar for all vowel pairs and for all speakers. To obtain an objective measure, a one-factorial ANOVA was calculated with a following Scheffé post hoc test with F0 value at the vowel mid as dependent variable and vowel identity as independent

factor (split by speaker). Consistently, all speakers show significantly lower F0 values for the low vowels /ɑː a/ compared to the other vowels /iː ɪ uː ʊ/. Thus results for IF0 found in the literature are replicated: The low vowels have a significantly lower F0 compared to the high vowels /iː ɪ uː ʊ/. Further, no significant difference could be found between high vowels differing in frontness (iː ɪ versus uː ʊ). Comparing the tense and lax vowel pairs in the ANOVA, for none of the speakers a significant difference could be found. Thus, the F0 between tense and lax vowels in the vowel mid is in fact similar in our data and replicate the results in the literature.
Extending the analysis to the landmarks onset and offset of the vowel, the vowel onset shows a similar significance pattern as the vowel mid. All high vowels are produced with significantly higher F0 values than low vowels. The tenseness difference is non-significant (vowel pair /uː ʊ/ for speakers JD shows differences). For the offset of the vowel, only the tense vowel /ɑː/ is significantly different from the other vowels (for speaker CG no significance). Comparing tense and lax vowel pairs, again no speaker shows significant differences.

3.2. Articulation

Assuming the tongue height to be the responsible articulator causing IF0, the vertical tongue back position would best capture the articulatory tongue height. Therefore, figure 2 shows boxplots for the vertical tongue back positions at the vowel mid points. As can be expected, all speakers show clear differences in tongue height between high and low vowels. To examine tongue height differences between the high and low vowels on the one hand and the tense and lax vowel pairs on the other hand, ANOVAs were calculated with the vertical positions of the tongue back and tongue blade sensors as dependent variables. At the vowel mid, obviously both sensors for the speakers CG, JD and RW show significant differences for the high versus low comparison. For the tenseness difference, the vowel pairs /iː ɪ/ and /uː ʊ/ showed significantly lower tongue height for the lax vowels. Additionally speaker CG had a significantly lower tongue position for tense /ɑː/ compared to lax /a/ which is typical for his South German dialectal background.

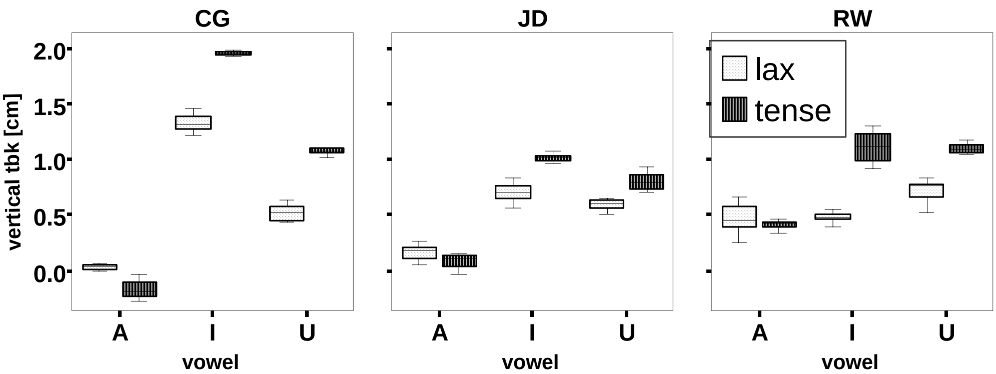

Figure 2: Boxplots of the vertical tongue back positions measured at the acoustically defined vowel mid, given separately for each speaker.

For the onset and offset of the vowel, exactly the same significance pattern for both tongue sensors was observed. Since the vertical jaw position naturally correlates with vowel height, in addition the significance pattern for this articulator was examined: The height difference was significant for all speakers. The tenseness difference values correspond to the significance values of the tongue sensors (difference: non-significant for vowel /ɑː/ versus /a/ for speaker RW, but significant for speaker JD). However, in a bite-block experiment Mooshammer et al. (2001) could exclude the jaw as the responsible articulator causing IF0 differences, thus it will not be regarded further in the following analyses.

3.3. Regression (Prediction model to predict F0 by the articulatory data)

With the aim to find the articulator which would predict most reliable the IF0 differences found in the F0 data of the different vowels, a stepwise regression model was computed. According to the biomechanical theories, the vertical tongue positions should give the most reliable explanation of the F0 variance, since these sensors capture best the tongue height which is said to be the factor causing the IF0 differences. Therefore, at the vowel midpoint F0 values were used as dependent variable and articulatory positions as independent variables (vertical and horizontal tongue and jaw positions) for a stepwise regression analysis. To compare our results to the data given in the literature the vowel mid was chosen, since IF0 values are mostly given at the vowel mid. Table 1 shows the extracted regression models with the predictors selected by the procedure, the explained variance (R^2), the F-values and the probability showing the significance of the model.

Speaker	Model	R^2	F	Probability
CG	1:TBKY 2: TBKY, TTPY	0.453 0.522	34.702 22.41	<0.001 <0.001
JD	1: TBKY 2: TBKY, TBLX	0.626 0.738	75.443 62.024	<0.001 <0.001
RW	JAWY	0.459	39.01	<0.001

Table 1: Regression models computed by the procedure linear stepwise regression with the selected articulator for the model. Further, the explained variance (R^2) is given and the F-values and the probabilities of the corresponding test for significance.

As can be seen, for two speakers (CG and JD) the vertical tongue back position was selected as the only relevant articulator. For speaker CG an additional model was computed with the inclusion of the vertical tongue tip sensor. However, due to the high correlation with the other variable tongue back sensor (0.88**), the tongue tip does not seem to contribute directly to the model. For speaker JD in the computation of the second model, the variable horizontal tongue blade sensor can be regarded as a suppressor variable, since it correlates only weakly with the dependent variable (-0.30) but highly with the other independent variable vertical tongue back sensor (0.80**). Thus the suppressor variable does not improve the model for the given analysis. For speaker RW, only the vertical jaw position was the variable that met the criterion for inclusion and explains 45.9% of the variance, but note that the correlation between the vertical tongue back and F0 is also very high (0.60**), although not as high as between the jaw and F0 (0.68**). Due to the higher correlation no model with the vertical tongue back position is computed, since it would not improve the explained variance further, but it can be regarded to be important in the interpretation of the results. In summary, it is shown that the vertical tongue back position is the articulator which explains the largest amount of variance in the F0 data for all speakers. The amount of explained variance is between 45% and 62%.

3.4. Correlation and regression of articulatory data and F0 comparing tense and lax vowels

Given the biomechanical theory is valid with regard to the German tense versus lax vowel pairs, the high/closed tense vowels should show a higher F0 due to their higher tongue positions compared to their lax counterparts (given that no additional enhancement takes place). Thus, in this case a correlation between acoustic F0 and articulatory positions pooled over each tense and lax vowel pair should be high and significant. In contrast, if the speaker uses an active enhancement strategy (as proposed by Hoole et al. 2004), the correlation should be weak and non-significant.
Thus, we examined these correlations with the aim to test whether the physiological linkage hypothesis between articulation and F0 is valid for German tense versus lax vowels. For the articulatory data, the vertical position of the tongue back sensor was chosen, since the stepwise regression analyses selected only this articulator (see 3.3). Additionally to the landmark vowel mid, correlations were also computed for the landmarks vowel onset and vowel offset in order to investigate whether the relationship F0-articulation would be different at these time-points. Thus, table 2 gives the correlation between the F0 values and the vertical tongue back sensor data for (1) all tense and all lax vowels (and for all vowels for comparison) and (2) for each of the three pooled tense and lax vowel pairs.

	speaker	All	All tense	All lax	/iː/ and /ɪ/	/uː/ and /ʊ/	/ɑː/ and /a/
Onset	**CG**	**.52 (0.27)**	**.62 (0.39)**	**.52 (0.28)**	.18 (0.00)	.031 (0.03)	.02 (0.00)
	JD	**.68 (0.47)**	**.62 (0.38)**	**.81 (0.65)**	.26 (0.25)	**.56** (0.07)	**.50 (0.32)**
	RW	**.68 (0.46)**	**.85 (0.72)**	.25 (**0.06**)	**.49** (0.34)	**.54** (0.24)	**.58 (0.29)**
Mid	**CG**	**.66 (0.66)**	**.62 (0.35)**	**.78 (0.60)**	.11 (0.15)	.16 (0.00)	.09 (0.25)
	JD	**.79 (0.63)**	**.78 (0.58)**	**.81 (0.65)**	.01 (0.27)	**.51** (0.12)	**.61** (0.13)
	RW	**.60 (0.36)**	**.82 (0.67)**	.12 (0.04)	.10 (0.15)	.37 (0.12)	.12 (0.12)
Offset	**CG**	**.32 (0.10)**	**.33 (0.11)**	.24 (0.06)	.36 (0.03)	.25 (0.12)	-.17 (0.06)
	JD	**.74 (0.55)**	**.76 (0.57)**	**.78 (0.60)**	.20 (**0.42**)	.15 (0.04)	**.65** (0.02)
	RW	**.39 (0.15)**	**.79 (0.62)**	.01 (0.00)	.12 (0.10)	-.15 (0.02)	.32 (0.01)

Table 2: Correlation coefficients between F0 values and vertical tongue back positions at the vowel onset, mid and offset. Correlations were (1) pooled for all vowels /iː ɪ uː ʊ ɑː a/, for all tense vowels /iː uː ɑː/ and all lax vowels /ɪ ʊ a/ and (2) pooled for the vowel pairs /iː ɪ/, /uː ʊ/ and /ɑː a/, all computed separately for each speaker.
In brackets, the explained variance and significance of linear regression models for the regression of the articulatory vertical tongue back position on the corresponding F0 value is given (significant values at $p<0.05$ are printed in bold).

As expected, the high correlation found for all vowels and for pooling all tense and lax vowels gives further evidence to the connection between the F0 values and the tongue position. Examining the tense and lax vowel pairs at the vowel mid, it can be seen that only one speaker (JD) shows a high and significant correlation for the vowels /uː ʊ/ and /ɑː a/, all other correlations are weak and non-significant. For the landmarks vowel onset and offset, the same pattern with weak correlations can be observed (additionally, speaker RW shows high correlations at the onset of the vowel). Thus, all these results seem to strengthen the idea of an active enhancement: Without this enhancement, the correlation would be high for the tense and lax vowel pairs, but due to the non-differing F0 values the correlation is found to be weak.
To see to which amount the vertical tongue sensor contributes to the variation of the F0 data, for each speaker linear regression models were computed examining the regression of the articulatory position onto the F0 values. The explained variance is given in brackets in table 2 following the correlation coefficient. As can be seen, the results are comparable to the correlation data: High and strongly significant explained variance for all vowels and the pooling of tense and lax vowels is given, whereas the explained variance is low and non-significant for the comparison of each tense and lax vowel pairs. The results are similar for the examination of the different acoustical landmarks vowel onset, mid and offset.

4. Conclusion

Results of the stepwise regression procedure and correlation tables give further evidence that the most important articulatory movements causing IF0 differences are the vertical tongue and the vertical jaw gestures. Examining the difference between German tense and lax vowels, the presented results strengthen the finding that the assumed higher tongue pull for the tense vowels in comparison to the lax vowels indeed does not induce a higher F0. It was given evidence by both correlation and regression analyses that the discrepancy persists between articulatory and F0 data, indicated by the weak and non-significant results when examining the tenseness distinction. Since the biomechanical link cannot explain these results, obviously other mechanisms seem to play a crucial role. As was given evidence by Hoole et al. (2004), speakers seem to actively increase F0 for lax vowels. However, the reasons for this active increase are not clear. One hypothesis would be that the perceptual cue vowel pitch (extracted by F0) is used to robustly identify tense versus lax vowels in noisy and/or ambiguous conditions: Due to the biomechanical link theory, the listener is expecting a lower F0 than actually present in the acoustic signal of the lax vowels. Thus this difference could aid in the identification in difficult acoustic conditions. For example, these conditions could include background noise masking certain formant values, ambiguous formant values caused by increased speaking rate or coarticulatory effects. Obviously, in all these cases the listener is faced with ambiguous main cues (for example similar formant values for tense and lax vowels: i.e. /eː/ vs. /ɪ/) to identify the vowel in question. Therefore, the cue vowel pitch could be additionally used to robustly enhance discriminability between the vowel classes. To test this hypothesis in order to see if the effect is perceptually important and valid, we conducted vowel pitch perception experiments with German speakers (Pape et al., 2005). However, the results gave evidence that vowel pitch is not consistently used for German tense and lax vowel pairs as a perceptual cue. Furthermore, the effect of Intrinsic pitch (the perceptual equivalent of IF0) is shown to be rather weak and variable across speakers. Thus, we could not give evidence that the active increase of F0 for the German lax vowels by the speakers is in fact aimed to increase perceptual validity of the different vowel classes for the listener. Since the perceptual validity of the active increase is questionable and the biomechancial link is shown to be insufficient to explain the differences, other reasons might be important for the higher F0 of lax vowels in German, i.e. other mechanisms causing the IF0 differences in general. However, these mechanisms could not be captured by our experimental setup.

As an outlook to further examinations, we aim to a better understanding of the relationship between articulatory movements and F0 and their interaction for the complete contours along the time axis. Both articulation and F0 movements are continuous movements with strong interaction, thus it is obvious that the chosen landmarks vowel onset, mid and offset only provide a small sample of the whole interaction. Thus, to allow an examination of both the complete sets of movements and repetitions along the time axis, a different approach compared to traditional statistics has to be regarded. Functional data analyses was proven to allow the examination in all these degrees of freedom and will be used for further analyses.

References

Diehl R.L. and Kluender. „On the objects of speech perception”, *Ecol. Psychol.*, 1, 121-144, 1989.

Fischer Jørgensen E. Intrinsic f0 in tense and lax vowels with special reference to German., *Phonetica* 47:99-140, 1990.

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, France, 2004.

Mooshammer C, Hoole P., Alfonso P. and Fuchs S. “Intrinsic Pitch in German: A Puzzle?”, *142nd Meeting of the ASA*, Ft. Lauderdale, Florida, USA, 2001.

Ohala J.J. and Eukel B.W. Explaining the intrinsic pitch of vowels, in *Honor of Ilse Lehsite*, edited by R.Channon and L. Shockey, Foris, Dordrecht, 207-215, 1987.

Pape D., Mooshammer C., Fuchs S. and Hoole, P. (2005). "Intrinsic pitch differences between German vowels /iː/, /ɪ/ and /yː/ in a cross-linguistic perception experiment", Proceedings ISCA workshop on plasticity in perception (PSP2005), London, UK.

Traunmüller H. “Perceptual dimension of openness in vowels”, *Journal of the Acoustical Society of America*, vol.69, pp. 1465-1475, 1981.

Whalen, D. and Levitt, A. (1995). “The universality of intrinsic F0 of vowels”, *Journal of Phonetics*, 23, 349-366.