From: In honor of Ilse Lehiste. Ed. by R. Channon & L. Shockey. Dordrecht: Foris. 1987

Explaining the Intrinsic Pitch of Vowels

John J. Ohala, *University of California, Berkeley, CA., U.S.A.*
Brian W. Eukel, *University of California, Berkeley, CA., U.S.A.*

1. INTRODUCTION

It has been known for several decades that, other things being equal, vowels have 'intrinsic' pitch (fundamental frequency, F_o) (Crandall 1925, Peterson and Barney 1952, Lehiste and Peterson 1961).[1] This is decidedly a second-order effect: it is small (high vowels have from 4 to 25 Hz higher pitch than low vowels) and acoustic theory shows that otherwise the sound source (the vibration of the vocal cords) is independent of the resonances of the vocal tract (Fant 1960). Nevertheless, an understanding of this phenomenon may lead to a better understanding of speech production in general and of the nature of vocal cord vibration and the ways F_o may be varied in particular.

Various hypotheses have been proposed to account for this effect (henceforth called 'intrinsic pitch') and of these, two receive serious consideration in the current literature: the 'coupling' hypothesis and the 'tongue-pull' hypothesis (see Ohala 1973, Atkinson 1973, Delos, Guérin, Mrayati, and Carré 1976 for reviews of this issue).

The coupling hypothesis, advocated by Lieberman (1970) and by Atkinson, suggests that when the first formant of the vowel is near the F_o of voice, as is the case with high vowels, acoustic coupling occurs between the vocal tract and the vocal cords similar to that which happens between the resonances of the bugle and the bugler's lips. In other words, the resonances of the vocal tract effectively 'dictate' to the vocal cords the frequencies at which they can vibrate (Ishizaka and Flanagan 1972, Benade 1973).

The tongue-pull hypothesis, offered by, among others, Ladefoged (1964*a*, *b*; but see below) and Lehiste (1970), states that somehow the tongue, in producing high vowels, pulls on the larynx and thus increases the tension of the vocal cords and thus the pitch of voice.

There are problems with both hypotheses.

Although it is easy enough to demonstrate source-tract coupling when phonating into a long tube (> 1 meter), when the impedance of the air mass in the resonating cavity approaches or exceeds the impedance of the source (Ishizaka and Flanagan), no one has found a way to determine directly the magnitude of the effect in normal speech. Accordingly, the evidence in favor of the coupling hypothesis comes exclusively from models of the vocal tract and the vocal cords. Some models exhibit coupling of sufficient magnitude to account for intrinsic pitch (Flanagan and Landgraf 1968) and some don't (Ishizaka and Flanagan). Some show the magnitude of the effect to be very small and in the *reverse* direction to what would be predicted to account for intrinsic pitch (Guérin 1978). Ohala (1973) pointed out that there is an opportunity to test the coupling hypothesis by examining the effect on intrinsic pitch when vowels are uttered in a helium-air mixture such that the resonances of the vocal tract are systematically raised, that is, increasing the separation between the first formant and F_o. (Helium-air has no effect or inconsistent effects on F_o itself.) Under such circumstances one would expect the differences in intrinsic pitch between high and low vowels to be lessened if coupling caused the intrinsic pitch. In fact, Beil (1962) showed that the correlation between vowel and mean F_o is largely unchanged when speaking in a helium-air mixture.

Other, more recent studies also cast doubt on the coupling hypothesis. Ewan (1979) found the intrinsic pitch effect during the labial nasal consonant /m/ preceding high and low vowels. The resonances of /m/ before high and low vowels are not thought to differ so much that they could lead to different amounts of coupling between vocal tract and vocal cords; on the other hand, it is quite plausible that there would be some anticipatory coarticulation during the /m/ of the tongue configuration of the following vowel. Similarly, Reinholt Petersen (1980) found that the intrinsic pitch of V_2 in a V_1CV_2 utterance was significantly influenced by the quality (specifically, the height) of V_1 even though there was an insignificant influence of V_1 on V_2's resonances. Given certain assumptions about how tongue-pull might operate, he interpreted these data as supporting that hypothesis over the coupling hypothesis.

The principal problem with the tongue-pull hypothesis is that it is far from clear how, exactly, the tongue may exert tension on the vocal cords in such a way as to raise the F_o of voice (but see below). Moreover, some of the anatomical structures whose displacement one would expect to be positively correlated with intrinsic pitch do not show this correlation. For example, from the phonetic literature one can find two basic patterns of how the vertical displacement of the larynx-hyoid bone complex varies as a function of vowel quality. Ranking the vowels according to the vertical position of the larynx-hyoid bone complex during their production, from high to low, one pattern is [i, ɑ, u] (Scripture 1902, Parmenter, Treviño, and Bevans 1933, Amenomori 1960, 1961, Shimizu 1960, 1961, Sundberg 1969, Ewan and Krones 1974) and a second pattern is [ɑ, i ≅ u][2] (Chiba and Kajiyama 1958, Öhman 1967, Houde 1968, MacNeilage 1969, Perkell 1969, Ladefoged, DeClerk, Lindau, and Papçun 1972). These facts led Ladefoged to abandon the tongue-pull hypothesis (Ladefoged et al.).

One might think of saving the tongue-pull hypothesis in the face of this damaging evidence by denying that there need be any correlation between larynx height and pitch, but this escape route seems to be blocked by the rather extensive evidence that, other things being equal, the vertical position of the larynx *is* positively correlated with F_o (Herries 1773, Magendie 1824:139, Wyllie 1866, Scripture 1902, Inoue 1931, Ohala 1978; see Ohala 1973 for further references).

Nevertheless, there may still be hope for the tongue-pull hypothesis. As mentioned, larynx height correlates with F_o, *other things being equal*; when looking at how the position of the larynx varies as a function of vowel quality, other things are not equal, viz., the configuration of the tongue. There may be other ways that the tongue could exert tension on the vocal cords that are independent of whatever mechanisms require the larynx-hyoid complex to move with pitch and with vowel quality. Ohala (1972, 1977, 1978) has suggested that the tissues connecting the tongue dorsum with the ary-epiglottic folds, which, in turn, are connected with the false vocal cords, could exert an indirect pull on the vocal cords either by inducing vertical tension in the vocal cords or by enlarging the laryngeal ventricle and in that way reducing the damping of the false cords on the vibration of the vocal cords proper. In fact, van den Berg (1955) and Shimizu (1961) present data showing that the ventricle is enlarged during the production of high vowels vis-a-vis low vowels. (That ventricle enlargement has *something* to do with increase of F_o is shown by Tsuboi 1959, Arnold 1961, and Keenan and Barret 1962.)

Other possibilities exist. As a variant of this scheme, Ewan (1979) proposes that the pharyngeal constriction characteristic of low back vowels such as [ɑ] must compress the supralaryngeal tissues and in that way add to the effective mass of the vocal cords, thus lowering F_o. In support of this, he found that pharyngeal consonants in Arabic, which have a marked retraction of the tongue dorsum, also have low intrinsic pitch. Honda and Baer (1981) argue that a contraction of the geniohyoid muscle (to bunch and raise the tongue for the production of high vowels) pulls the hyoid bone forward and in that way tilts the thyroid cartilage down in such a way as to augment the anterior-posterior tension of the vocal cords.

Consistent with all three of these mechanisms (and therefore weakly supportive of the tongue-pull hypothesis) are observations attesting to systematic differences in the physical configuration of the laryngeal vestibule and/or structures in the immediate vicinity (e.g., the epiglottis) in ways which put high vowels and low vowels at opposite ends of a scale (Eijkman 1933, Russell 1928:89-132). Eijkman, for example, orders the vowels according to the degree of tilt of the epiglottis during their production, from most forward to most backed, as follows: [i, e, u, o, ɛ, a, ɑ, ɔ].

It must be acknowledged, however, that the evidence just reviewed as being consistent with the tongue-pull hypothesis is only circumstantial; a correlation between certain articulatory measures and intrinsic pitch doesn't guarantee that there is any causal nature to the correlation. What is needed is a way to vary the pull of the tongue on the larynx without any concomitant changes in the resonances of the vocal tract. If the intrinsic differences in the pitches of vowels are enhanced when tongue pull is increased it would be strong support for the tongue-pull hypothesis. We report here an attempt to implement such an experiment. The procedure was suggested to us by Björn Lindblom (personal communication).

2. METHOD

Lindblom and Sundberg (1971) showed that speakers can produce acoustically normal vowels when speaking with the jaw in abnormal fixed positions if they produce compensatory tongue gestures. For example, by increasing the elevation of the tongue they can produce good [i, u] vowels with their jaws propped open by bite blocks. The tracings they provide of the sagittal profiles of the tongue under these conditions suggest that the 'stretch' between the tongue and the larynx is greater than normal. On the other hand, speaking low, open vowels [æ, ɑ] with a propped open jaw approximates normal conditions and thus should not appreciably change the degree of stretch between the tongue and larynx. Briefly, then, the purpose of the experiment was to see if the magnitude of the *difference* between the intrinsic pitch of high vowels and low vowels was increased when subjects spoke with bite blocks propping their jaw open. This is represented formally as follows:

Hypothesis: $(\text{Max int. } F_o - \text{Min int. } F_o)_{\text{bite block}} >$
$(\text{Max int. } F_o - \text{Min int. } F_o)_{\text{normal}}$

The null hypothesis, of course, would be that the bite block condition would not increase the intrinsic pitch differences between vowels.

We therefore had 8 adult speakers of American English, 4 males and 4 females, pronounce English nonsense words of the type [ə'mimə], [ə'mɛmə], etc., where the middle (stressed) vowel was any of the monophthongal vowels [i, u, ɛ, ʌ, æ, ɑ], approximately 16 tokens each, under four conditions: normal jaw mobility, jaw propped open with small wooden spacers 1 mm., 5 mm., and 10 mm. thick, placed bilaterally between the molars. The order of the conditions was varied for each subject so that order effects would cancel. Speakers were given frequent 5 or 10 minute breaks during the readings. Three speakers read the randomized sequence of words from a printed list but because they had some difficulty figuring out the intended phonetic quality of the medial vowel from the way they were spelled, the remaining 5 subjects spoke the words as they heard them from a cue tape.

All speech samples were recorded on a high quality tape recorder and later played back through a custom-made analog pitch extractor and written out

oscillographically. The peak pitch of each token was measured by hand. Each speaker's pitch values were normalized by replacing all values by their corresponding z-score, i.e., the deviation of the given sample from the mean divided by the standard deviation, the mean and standard deviation computed from all utterances of the given speaker over all four conditions. Normalization was deemed necessary so that we could meaningfully average pitch data from speakers with different overall pitch levels.

The data from one female subject had to be excluded from the study since half way through the session her pitch level dropped about 10 to 15 Hz and thus did not lend itself to the simple normalization process we used. Also discarded were a few tokens in which the stressed vowel was unlike any of the six target vowels.

The pattern of data from the male and female speakers was similar enough to justify pooling all data for the final analysis.

Auditory analysis of all tokens and an acoustic analysis of a small random sample of the vowels showed, in agreement with the findings of Lindblom and Sundberg, that there was no consistent acoustic difference between the vowels spoken with and without bite blocks.

3. RESULTS

The normalized pitch measurements for the 7 subjects are given in Table 1. The columns separate the data according to the vowel, the rows, according to

Table 1. Values for the mean normalized pitch, the standard deviation (in parentheses), and the number of observations for the indicated vowel (columns) in the indicated condition (rows). Data drawn from 7 subjects.

Vowel / Condition	ɑ	æ	ɛ	ʌ	i	u
Normal	−0.81 (0.77) 99	−0.58 (0.87) 115	−0.65 (0.72) 111	−0.53 (0.99) 118	0.23 (0.86) 114	0.44 (0.82) 116
1 mm.	−0.18 (0.74) 109	−0.33 (0.82) 103	−0.15 (0.81) 110	−0.21 (0.78) 109	0.72 (0.82) 106	0.93 (0.77) 113
5 mm.	−0.11 (0.87) 115	−0.25 (0.81) 110	−0.17 (0.76) 108	−0.14 (0.71) 115	0.78 (0.89) 103	0.93 (0.91) 110
10 mm.	−0.47 (0.79) 111	−0.47 (0.79) 119	−0.29 (0.90) 103	−0.24 (0.86) 107	0.62 (0.97) 114	1.09 (1.00) 108

the experimental condition. The entries in each cell give the mean normalized pitch, the standard deviation (in parentheses), and the number of observations these first two measures are based on.

4. DISCUSSION

In accord with previous studies, high vowels had higher pitch than low vowels in all conditions, with the 'mid' vowels [ɛ, ʌ] patterning with the low vowels. The difference between maximum and minimum pitches is significant in all cases. This difference is greater in the 10 mm. bite block condition (1.09 -(-0.47) = 1.56) than in the normal condition (0.81 - (-0.44) = 1.25) and this difference is significant ($p < 0.05$, $t = 1.84$, degrees of freedom = 205, one-tailed t-test).[3] These data therefore support the original hypothesis.

There is no significant difference between the maximum – minimum pitch intervals in the normal vs. either the 1 mm. or 5 mm. conditions. This suggests that these smaller bite blocks did not require sufficient compensatory adjustment of the tongue configuration in a way that would augment the tongue pull. They also suggest that it is not simply the artificial experimental condition of speaking with something clenched between the teeth which gives rise to the enhanced pitch differences.

Lubker, McAllister, and Lindblom (1977) performed a replication of this experiment using Swedish speakers and reported negative results. However, they did not normalize their subjects' pitch data as we did, so the data are not comparable.

5. CONCLUSION

The purpose of this study was to devise a test which would differentiate between two competing hypotheses on the cause of the intrinsic pitch differences of vowels. Accordingly we created a circumstance where one of the hypothesized causal factors, vocal tract resonances, was held constant, while the other causal factor, degree of tongue pull on the larynx, was varied. Previous research suggested that speaking high vowels with the jaw artificially propped open would increase the tension between tongue and larynx without affecting vowel quality appreciably. The results showed that the difference in intrinsic pitch between high and low vowels was increased under the condition of speaking with propped-open jaw, thus supporting the hypothesis that tongue pull is responsible for the intrinsic pitch of vowels. It still remains to be determined, though, exactly *how* the pull of the tongue can influence vocal cord tension which in turn determines F_o.

ACKNOWLEDGEMENTS

This research was supported by a grant from the National Science Foundation.

NOTES

1. This is a revised version of a paper originally presented at the 92nd Meeting of the Acoustical Society of America, San Diego, 15-19 November 1976, and which appeared in *Report of the Phonology Laboratory* (Berkeley), 1978, No. 2, pp. 118-125.
2. Houde offered a plausible explanation for this second pattern where low vowels have a higher hyoid bone position than high vowels: A contraction of the hyoglossus muscle is required to lower and flatten the tongue for [ɑ] but would be absent for high vowels. A contraction of this muscle would be expected to raise the hyoid bone, to which it is anchored; the hyoid bone itself is only loosely connected to structures below it.
3. For the purposes of this t-test, the variance of the pitch difference in a given condition was set equal to the sum of the variances of the minimum and maximum pitch levels in each condition since independent variances are additive. *N*, the population, was set as the lesser of the two *N*'s involved in deriving the pitch difference.

REFERENCES

Amenomori, Y. (1960). Significance of the hyoid bone in phonation (Analysis of the suspension mechanism of the larynx). (In Japanese.) *Oto-rhino-laryngology Clinic (Kyoto)* **53**, 117-144.

Amenomori, Y. (1961). Some studies on the 'suspending mechanism of the larynx.' (In Japanese.) *Studia Phonologica* **1**, 95-110.

Arnold, G.E. (1961). Physiology and pathology of the cricothyroid muscle. *Laryngoscope* **71**, 687-753.

Atkinson, J.E. (1973). *Aspects of intonation in speech: implications from an experimental study of fundamental frequency.* Doc. diss., University of Connecticut.

Beil, R.G. (1962). Frequency analysis of vowels produced in a helium-rich atmosphere. *J. Acous. Soc. Am.* **34**, 347-349.

Benade, A. (1973). The physics of brasses. *Scientific American*, July, pp. 24-35.

Berg, J.W. van den (1955). On the role of the laryngeal ventricle in voice production. *Fol. Phoniatrica* **7**, 57-69.

Chiba, T. and Kajiyama, M. (1958). The vowel, its nature and structure. *Phonetic Soc. of Japan.*

Crandall, J.B. (1925). The sounds of speech. *Bell Sys. Tech. J.* **4**, 586-626.

Delos, M., Guérin, B., Mrayati, M., and Carré, R. (1976). Study of intrinsic pitch of vowels. *J. Acous. Soc. Am.* **29**, S72.

Eijkman, L.P.H. (1933). The internal aspect of the larynx in speech. *Archives Néer. Phonétique Expér.* **8-9**, 213-219.

Ewan, W.G. (1979). Can intrinsic vowel F_o be explained by source/tract coupling? *J. Acous. Soc. Am.* **66**, 358-362.

Ewan, W.G. and Krones, R. (1974). Measuring larynx movement using the thyroumbrometer. *J. Phonetics* **2**, 327-335.

Fant, G. (1960). *Acoustic theory of speech production.* The Hague: Mouton.

Flanagan, J.L. and Landgraf, L. (1968). Self oscillating source for vocal-tract synthesizers. *IEEE Trans. on Audio and Electroacoustics* **AU-16**, 57-64.

Guérin, B. (1978). *Contribution aux recherches sur la production de la parole; Etude du fonctionnement de la source vocale; Simulation d'un modele*. Doc. diss., L'université Sciéntifique et Médicale de Grenoble.

Herries, J. (1773). *The elements of speech*. London. (Reprinted 1968 by the Scholar Press, Ltd., Menston.)

Honda, K. and Baer, T. (1981). Interaction between pitch control and articulatory mechanisms. *J. Acous. Soc. Am.* **70**, S12.

Houde, R.A. (1968). *A study of tongue body motion during selected speech sounds*. Speech Communication Research Laboratory, Inc. (Santa Barbara) Monog. No. **2**.

Inoue, O. (1931). Two methods of pitch recording devices. (In Japanese.) *Study of Sounds* **4**, 57-60.

Ishizaka, K. and Flanagan, J.L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Sys. Tech. J.* **51**, 1233-1268.

Keenan, J.S. and Barret, G.C. (1962). Interlaryngeal relationships during pitch and intensity changes. *J. Speech and Hearing Res.* **5**, 173-178.

Ladefoged, P. (1964*a*). Some possibilities in speech synthesis. *Lang. and Speech* **7**, 205-214.

Ladefoged, P. (1964*b*). *A phonetic study of West African languages. An auditory-instrumental survey*. Cambridge Univ. Press.

Ladefoged, P., DeClerk, J., Lindau, M., and Papçun, G. (1972). An auditory-motor theory of speech production. *Working Papers in Phonetics (UCLA)* **22**, 48-75.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge: M.I.T. Press.

Lehiste, I. and Peterson, G. (1961). Some basic considerations in the analysis of intonation. *J. Acous. Soc. Am.* **33**, 419-425.

Lieberman, P. (1970). *A study of prosodic features*. Status Rep. on Speech Res., Haskins Laboratories. SR-23-1970, 179-208.

Lindblom, B. and Sundberg, J. (1971). Neurophysiological representation of speech sounds. *Proc., 15th World Cong. of Logopedics and Phoniatrics, Buenos Aires*.

Lubker, J., McAllister, R., and Lindblom, B. (1977). Vowel fundamental frequency and tongue height. *J. Acous. Soc. Am.* **62**, S16-S17.

MacNeilage, P. (1969). A note on the relation between tongue elevation and glottal elevation in vowels. *Monthly Internal Memorandum (Berkeley), January*, pp. 86-90.

Magendie, F. (1824). *An elementary compendium of physiology*. (Transl. by E. Milligan.) Philadelphia.

Ohala, J.J. (1972). How is pitch lowered? *J. Acous. Soc. Am.* **52**, 124.

Ohala, J.J. (1973). Explanations for the intrinsic pitch of vowels. *Monthly Internal Memorandum (Berkeley)*, January, pp. 9-26.

Ohala, J.J. (1977). Speculations on pitch regulation. *Phonetica* **34**, 310-312.

Ohala, J.J. (1978). The production of tone. In: V.A. Fromkin (ed.), *Tone: a linguistic survey*. New York: Academic Press. 5-39.

Öhman, S. (1967). Numerical model of coarticulation. *J. Acous. Soc. Am.* **41**, 310-320.

Parmenter, C.E., Trevino, S.N., and Bevans, C.A. (1933). The influence of a change in pitch on the articulation of a vowel. *Lang.* **9**, 72-81.

Perkell, J.S. (1969). *Physiology of speech production. Results and implications of a quantitative cineradiographic study*. Cambridge: M.I.T. Press.

Peterson, G.E. and Barney, H.L. (1952). Control methods used in a study of the vowels. *J. Acous. Soc. Am.* **24**, 175-184.

Reinholt Petersen, N. (1980). Coarticulation of inherent fundamental frequency levels between syllables. *Ann. Rep., Inst. of Phonetics, Univ. of Copenhagen* **14**, 317-354.

Russell, G.O. (1928). *The vowel, its physiological mechanism as shown by x-ray*. Columbus: The Ohio State Univ. Press.

Scripture, E.W. (1902). *The elements of experimental phonetics*. New York.

Shimizu, K. (1960). On the motions of the vocal cords in phonation studied by means of the high voltage radiograph movies. (In Japanese.) *The Oto-rhino-laryngology Clinic (Kyoto)* **53**, 446-461.

Shimizu, K. (1961). Experimental studies on movements of the vocal cords during phonation by high voltage radiograph motion pictures. *Studia Phonologica* **1**, 111-116.

Sundberg, J. (1969). Articulatory differences between spoken and sung vowels in singers. *Quart. Prog. and Status Rep., Speech Transmission Lab., Royal Inst. of Tech., Stockholm*. STL QPSR 1/1969, 33-46.

Tsuboi, H. (1959). Morphologic changes of the larynx and inferior resonant cavity in phonation, studied by means of high-voltage radiography. (In Japanese.) *J. of the Oto-rhino-laryngological Soc. of Japan* **62**, 2064-2085.

Wyllie, J. (1866). Observations on the physiology of the larynx. *Edinburgh Med. J.*, September, pp. 214-237.