Integrating Voice Quality and Tongue Root Position in Perceiving Vowels

John Kingston, Laura Walsh, Rachel Thorburn, and Christine Bartels
Linguistics Department, South College, University of Massachusetts, Amherst
Neil A. Macmillan
Psychology Department, Brooklyn College of the City University of New York

Abstract

In two experiments, we examined the perceptual interactions between the acoustic correlates of covarying articulations in vowels: tongue root advancement and tense-lax voice quality. These correlates prove to integrate because they both contribute to the perceived flatness or sharpness of a vowel's spectrum, and this integration enhances the contrast between vowels.

Introduction

Vowels articulated with the tongue root advanced are often also produced with a lax or breathy voice quality, whereas retracted tongue root vowels are produced with a tense or creaky voice quality; similar covariation is observed between tongue body raising and a lax voice quality [1]. The aryepiglottal ligament and membrane connect the tongue root to the arytenoid cartilages and may cause them to slide forward slightly and/or rock slightly apart, slackening or separating the vocal folds enough to lax the voice, when the tongue root is advanced or the body raised. However, because some languages, e.g. Dinka, exhibit independent contrasts for tongue root/body position and voice quality, it appears there is no necessary physiological interaction between the lingual and laryngeal articulations.

The experiments reported here assess an alternative, perceptual explanation for their covariation: that because lax or breathy voice causes energy to fall off more rapidly with increasing frequency in the source spectrum and advancing the tongue root lowers F_1, these articulations combine to bias a vowel's spectrum toward low frequencies, i.e. make it *flatter*, whereas a tenser or creakier voice quality and a non-advanced or retracted tongue root bias the vowel's spectrum toward higher frequencies, i.e. make it *sharper*. The covariation thus enhances the vowels' contrast along the *flat:sharp* dimension [cf. 2]. This explanation does not rely on a mechanical, physiological connection between the articulations that affect tongue root position (or body height) and the state of the vocal folds, but instead allows these articulations to be independently controlled by speakers. Kingston & Diehl [3, also 4] argue that speakers exert themselves to produce multiple articulatory differences between minimally contrasting phonemes when those articulations' acoustic correlates are similar enough psychoacoustically to integrate into higher-level perceptual properties as the flatness property proposed here.

Methods

Tense-lax variation in *V(oice) Q(uality)* was achieved by manipulating the percent of the glottal cycle in which the glottis is open and the additional energy reduction at 3 kHz in the source spectrum beyond the default decay of -6 dB/octave, i.e. the *O(pen) Q(uotient)* and *S(pectral) T(ilt)* parameters in the KLSYN88 terminal analogue synthesizer [5]. Variation in tongue root position was implemented simply through manipulation of F_1. The stimuli used in the two experiments reported here had similar ranges of F_1 values (Table 1), but they differed in what part of the voice quality continuum was paired with F_1: voice quality in the Tense experiment ranged from very tense to intermediate values along tense-lax continuum, whereas in the Lax experiment this dimension ranged from (overlapping) intermediate values to very lax. The overlap allows us to combine the results of the two experiments in constructing a map of how the entire range of voice qualities interacts perceptually with a narrower range of F_1s. Table 1 shows the 4x4 stimulus arrays defined by the orthogonal combination of F_1 and *VQ* values used in the two experiments. The size of the steps along the F_1 and *VQ* dimensions were approximately a jnd (at 70-80% correct). Other synthesis parameters were set so as to create a syllable of the shape [bVb], whose vowel was mid to high back in quality.

Table 1: Steady-state parameter values and 4x4 stimulus arrays for the Tense and Lax experiments: A-D = F_1 values (horizontal axis) and 1-4 and 5-8 = Ten(se) to Int(ermediate) and Int to Lax VQ values for the Tense and Lax experiments, respectively.

VQ			F_1			
			Advanced		Retracted	
	OQ	ST	470	484	499	514
Ten	29	-3	A1	B1	C1	D1
	33	-4	A2	B2	C2	D2
	39	-6	A3	B3	C3	D3
Int	53	-11	A4	B4	C4	D4
VQ			F_1			
	OQ	ST	450	468	506	536
Int	42	-7	A5	B5	C5	D5
	54	-11	A6	B6	C6	D6
	72	-17	A7	B7	C7	D7
Lax	90	-23	A8	B8	C8	D8

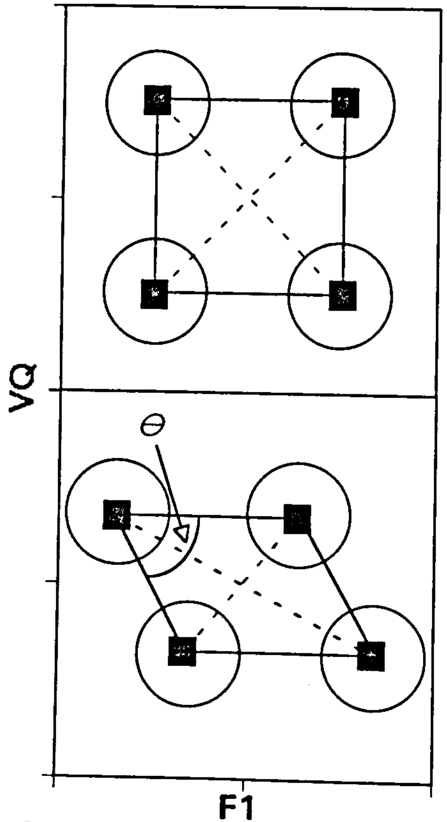
Two different groups of eight, paid, well-practiced, normal-hearing listeners participated in each experiment. Just a single stimulus was presented in each trial, to which the listener gave one of 2 responses, followed by a confidence judgment, and then by feedback. In the Lax experiment (run first) 16 alternating practice trials were followed by 96 randomized test trials; performance was assessed from the last 90 test trials/task/listener. In the Tense experiment, two blocks of 12 alternating practice trails followed by 66 randomized test trials were run for each condition, one early and one late. Performance was assessed from the last 60 test trials in each block, for a total of 120 trials/task/listener, an increase by one-third over the Lax experiment.

In the results reported here, listeners had to classify stimuli differing by one step in the 4x4 arrays in one of two ways: along just a single dimension or in a correlated fashion along both dimensions; these will be referred to as *single-dimension* and *correlated classification* tasks. There are 12 single-dimension classification tasks along each dimension, e.g. A1 vs A2 for *VQ* differences and A1 vs B1 for F_1 differences. In the correlated classification tasks, the correlation between F_1 and tense *VQ* could be either positive, e.g. A2 vs B1, or negative, e.g. A1 vs B2; there are 9 such tasks for each correlation polarity. Positive correlation corresponds to the natural covariation.

As in our previous work [6], we use detection theory [7] to model differences in how accurately our listeners classified the stimuli, because this theory provides the same estimate (d') of accuracy from each task. Furthermore, as d' estimates perceptual distance, it describes our listeners' performance as a mapping from the stimulus space defined by the dimensions F_1 and *VQ* to the two-dimensional *decision* space seen from above in Fig. 1, in which d's represent the distances between the mean locations of the stimuli (points in Fig. 1). The listener divides the decision space into response regions, but because various sources of noise produce trial-to-trial variability on both dimensions, the stimuli's perceptual values in the decision space form bivariate distributions of response likelihood (the circles in Fig. 1 suggest that the distributions are equal-variance, uncorrelated, bivariate normal).

We model our listeners' performance in terms of the d's obtained from 4 single-dimension and 2 correlated tasks that define each of the nine 2x2 subarrays of the larger 4x4 array in each experiment: the single-dimension d's determine the lengths of the sides of a parallelogram, the correlated d's the lengths of its diagonals. The rectangular arrangement in the upper panel of Fig. 1 is the arrangement of the stimuli when their perceptual values on one dimension do not depend on their values on the other, i.e. when the dimensions are *separable*. In the lower panel the means of the distributions are no longer arranged rectangularly, and the value of a stimulus on one dimension *does* depend on its value on the other; because the locations of the means of the distributions shifts in the space, we call the dimensions *mean-integral* in this case [6 cf. 9]. Comparing these two panels suggests that it is the (in)equality of d's obtained from the pair of correlated tasks representing the lengths of the diagonals in a 2x2 subarray that determines whether the stimulus dimensions are perceptually integral. The upper left angle (θ) of a parallelogram is a measure of mean-integrality: θ approaches 90° when the stimulus dimensions are perceptually separable but deviates from it when they're mean-integral, toward 0° when the negatively correlated task is easier than the positively correlated vs toward 180° when the positively correlated task is easier. The θ that provided the best-fitting parallelograms to the correlated and single-dimension d' values was found by iteration; the fit of the correlated to the single-dimension data tests the model. If F_1 and VQ integrate into the property called flatness above, then listeners should be uniformly more accurate on the negatively than the positively correlated tasks, i.e. $\theta < 90°$.

Figure 1: Parallelograms for separable and mean-integral dimensions.



Results

Table 2 lists the θs obtained for each 2x2 subarray from each experiment for each 2x2 subarray. The fits are clearly much better in the Tense than the Lax Experiment, probably because the number of trials contributing to each d' was increased by a third and because the listeners were more thoroughly pre-tested and trained in that experiment.

Table 2. θs (in degrees) for the parallelogram models of each 2x2 subarray of the 4x4 arrays in the Tense and Lax experiments (with rms errors in d' = standard deviation units).

VQ		*F_1*		
		A:B	B:C	C:D
Tense	1:2	49 (.113)	34 (.148)	52 (.129)
	2:3	84 (.299)	62 (.149)	43 (.087)
	3:4	107 (.544)	120 (.340)	127 (.174)
Lax	5:6	159 (.147)	120 (1.027)	180 (.620)
	6:7	0 (.830)	15 (.758)	27 (.785)
	7:8	56 (.344)	79 (.564)	19 (.245)

Both halves of Table 2 show evidence of strong mean-integrality effects, i.e. VQ and F_1, do interact; however, the extent and direction of the interaction, measured by θ, varies systematically with location in the arrays. For the 2x2 subarrays from the tenser row pairs 1 vs 2 and 2 vs 3 in the 4x4 array in the Tense experiment (top), that θ is clearly less than 90° shows the stimuli in which laxness and F_1 negatively covary are easier to classify. However, θs are clearly greater than 90° for the parallelograms of the laxest pair of rows, 3 vs 4, but θs also are not much greater than 90°, indicating that the dimensions may be near separable with intermediate voice qualities. However, for the 2x2s drawn from rows 5 vs 6 in the Lax experiment (bottom), whose voice qualities overlap with those in rows 3 vs 4 in the Tense experiment, θs are clearly all much greater than 90°, indicating strong mean-integrality in the opposite direction. With yet laxer voice qualities, the direction of mean-integrality flips once more, as θs are all obviously less than 90° for 2x2s drawn from rows 6 vs 7 and 7 vs 8. Thus, at the tense and lax ends but not the middle of the voice quality continuum, we find the direction of mean-integrality expected if the acoustic correlates of tongue root advancement and voice quality enhance the contrast between vowels by integrating into the perceptual property, flatness. In the middle, these dimensions may be separable or integrate strongly in the opposite direction, i.e. into a property of the spectrum we could call *compactness* (low F_1 and tense voice being compact in contrast to the diffuse combination of high F_1 and lax voice) rather than into flatness.

Discussion

The flips observed in the direction of mean-integrality may follow from simple psychoacoustic properties of the stimuli in these experiments. Because the laxer the voice the more rapidly source-spectrum energy falls off with increasing frequency and the more advanced the tongue root the lower F_1 is, the difference in amplitude between the first harmonic and those immediately above it should vary directly with flatness. To model these effects, we fit a line to the peaks of the first four harmonics for each stimulus. The slope of this line estimates how their relative amplitudes differ as a function of both voice quality and F_1: this slope should be more shallow or even more negative the laxer the voice and the lower the F_1, i.e. the flatter the vowel's spectrum. The difference in slopes between the stimuli in each of the correlated tasks measures how much they differ in flatness, and thus estimates the psychoacoustic value of flatness for predicting differences in relative accuracy on these tasks. When these slope differences were fit to the observed mean d's for the correlated tasks in a simple regression model, a positive change in d' of 0.37/dB difference in slope was obtained Although highly significant [$F(1,34) = 7.93, p = 0.008$], the proportion of variance accounted for was only 0.18. When the slope differences are fit only to the correlated data from the 2x2s at the tense and lax ends of the arrays, where θs were less than 90°, the relationship is more strongly positive, 0.53 change in d'/dB, and the proportion of variance accounted for jumps substantially, to 0.42 [$F(1,18) = 13.00, p = 0.002$]. The direction of mean-integrality at the two ends of the tense-lax may therefore arise from differences in the psychoacoustic property flatness, even if some other property is psychoacoustically more salient in the middle of this continuum.

[Work supported by NIH Grant R-29-DC01708 to first author and NSF grant DBS92-12043 to second author; earlier versions: Thorburn, R., *et al.* (1994), J. Acoust. Soc. Am. 95 (Abstract) and Walsh, L. *et al.* (1995), J Acoust. Soc. Am. 97 (Abstract).]

References

[1] Denning, K. (1989) *The diachronic development of phonological voice quality*, Ph.D. diss. Stanford.

[2] Jakobson, R., G. Fant & M. Halle. (1952), *Preliminaries to speech analysis*, Cambridge: MIT Press.

[3] Kingston, J. & R.L. Diehl. (1994), "Phonetic knowledge", *Lg.* 70: pp. 419-454.

[4] Diehl, R.L., J. Kingston, W.A. Castleman. (1995), "On the internal perceptual structure of phonological features: the [voice] distinction", *J. Acoust. Soc. Am.*, 97 (Abstract).

[5] Klatt, D.H. & L.C. Klatt. (1990), "Analysis, synthesis, and perception of voice quality variations among female and male talkers", *J. Acoustic. Soc. Am.* 87, pp. 820-857.

[6] Kingston, J. & N.A. Macmillan. (1995), "Integrality of nasalization and F_1 in vowels in isolation and before oral and nasal consonants", *J. Acoust. Soc. Am.* 97, pp. 1261-1285.

[7] Macmillan, N.A. & C.D. Creelman. (1991), *Detection theory: A user's guide*. Cambridge: Cambridge UP.