Prosodic Effects on Glottal Allophones

Janet B. Pierrehumbert

In O. Fujimura, and M. Hirano, eds)(1995) Vocal Fold Physiology 8: Voice Quality Control. Singular Press, San Diego. 39-60.

Singular Publishing Group, Inc.
4284 41st Street
San Diego, California 92105-1197

Typeset in 10/12 Stempel Schneidler Roman by So Cal Graphics
Printed in the United States of America by McNaughton & Gunn

Library of Congress Cataloging-in-Publication Data

Vocal Fold Physiology Conference (8th : 1994 : Kurume-shi, Japan)
Vocal fold physiology : voice quality control : proceedings of the VIIIth Vocal Fold Physiology Conference, Kurume, April 7–9, 1994 / edited by Osamu Fujimura, Minoru Hirano : with the assistance of Peter N. Ladefoged . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 1-56593-379-6
1. Vocal cords—Physiology—Congresses. 2. Voice—Congresses. I. Fukimura, Osamu, 1927– II. Hirano, Minoru, 1932– III. Title.
[DNLM: 1. Voice Quality—physiology—congresses. 2. Vocal Cords—physiology—congresses. 3. Voice Disorders—Physiopathology—congresses. WV 501 V872v 1994]
QP.V59 1994
612.7'8—dc20
DNLM/DLC
for Library of Congress 94-29735
CIP

Contents

vocal folds does not cause a large increase in tension. Therefore, there is no theoretical support for the vertical tension account.

Nevertheless, in some extreme conditions, like singing, this change of the laryngeal configuration in the vertical dimension may affect the tension of the vocal folds. The vertical tension applied on the hyoid-larynx complex may cause some F0 changes. For example, when the laryngeal ventricle is collapsed and the false vocal folds are in contact with the vocal folds, that should affect the vibratory mode of the vocal folds.

References

Colton, R. H., & Estill, J .A. (1981). Elements of voice quality: Perceptual, acoustic, and physiologic aspects. *Speech and Language: Advances in Basic Research and Practice, 5*, 311–403.

Hirai, H., Honda, K., Fujimoto, I., & Shimada, Y. (1994). Analysis of magnetic resonance images on the physiological mechanisms of fundamental frequency control [in Japanese]. *Journal of the Acoustic Society of Japan, 50*, 296–304.

Honda, K. (1983). Relationship between pitch control and vowel articulation. In D. M. Bless & J. H. Abbs (Eds.), *Vocal fold physiology* (pp. 286-297). San Diego: College-Hill Press.

Honda, K., Dang, J., & Hirai, H. (in press). A physiological model of speech organs and the implications of tongue-larynx interaction. In the *Proceeding of International Congress of Spoken Language Processing*, 1994.

Honda, K., Hirai, H., & Kusakawa, N. (1993). Modeling vocal tract organs based on MRI and EMG observations and its implication on brain function. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 27*, 37–49.

Kobayashi, N., & Honda, K. (1993). Production and perception of the consonantal voicing distinction in speech without VOT: An acoustic study using electric larynx speech. *ATR Technical Report*, TR-A-0172.

Ohala, J. J., & Eukel, B. W. (1978). Explaining the intrinsic pitch of vowels. *Report of the Phonology Laboratory, 2*, 118–125. Berkeley: University of California.

Sundberg, J. (1970). Formant structure and articulation of spoken and sung vowels. *Folia Phoniatrica, 22*, 28–48.

Stevens, K. N. (1981). Vibration modes in relation to model parameters. In K. N. Stevens & M. Hirano (Eds.), *Vocal fold physiology* (pp. 291-304). Tokyo: University of Tokyo Press.

Titze, I. (1988). A framework for the study of vocal registers. *Journal of Voice, 2*, 183–194.

Vilkman, E., & Karma, P. (1989). Vertical hyoid bone displacement and fundamental frequency of phonation. *Acta Otolaryngologica* (Stockholm), *108*, 142–151.

Yanagisawa, E., & Estill, J. A. (1989). The contribution of aryepiglottic constriction to "ringing" voice quality. *Journal of Voice, 3*, 342–350.

CHAPTER

3

Prosodic Effects on Glottal Allophones

Janet Pierrehumbert

1. INTRODUCTION

In some languages, a full glottal stop or a glottalized voice quality is phonologically contrastive. That is, words can be distinguished by the presence or absence of this property. This is not the case in English; nonetheless, glottalization and even full glottal stops are quite common. They arise from several different sources. Some speakers habitually produce a creaky voice quality. Others produce such a quality sporadically for rhetorical effect. Many speakers use a creaky voice quality as equivalent to super-low F0; that is, in places where the intonation pattern is canonically low, these speakers produce a creaky quality which, by virtue of the irregularity of the excitation, does not strictly speaking have an F0 value (see Anderson, 1986). However, speech with this type of excitation is perceptually similar to speech with an extremely low F0 and apparently counts as such.

This study concerns the fourth and most conventional source of glottalization in English, namely segmental allophony. Two allophony rules are relevant. First, a so-called glottal stop is often inserted at the beginning of a vowel initial word, especially if this word is either phrase ini-

tial or follows a word ending in a vowel. Some examples are given in example 1, where insertion sites are indicated with underscores.

(1) _Emma, _August is hot here.
You could fix the camp stove, the _oarlocks, the _icemaker . . .

Although a full glottal stop is possible at these locations, often one merely observes a disturbance of the voice quality. Second, voiceless stops, especially /t/, are frequently glottalized in the coda position of a syllable. This is a factor in the famous contrast in allophony between "nitrate" (in which the first /t/ is produced as an affricate) and "night-rate" (in which the first /t/ is glottalized and may even lose its coronal articulation altogether.) "Nitrate" is monomorphemic and is syllabified "ni.trate", with the /t/ parsed in an onset cluster with /r/ at the beginning of the second syllable. The syllabification of "night-rate", respecting the word boundary, is "night.rate"; the /t/ is in the coda of the first syllable. In both cases, the glottalization is believed to be produced by hyper-adduction of the vocal folds.

Because the topic of the paper is glottalization arising through segmental allophony, the data collection was specifically designed to eliminate the other sources of glottalization. The two speakers selected for the study have even modal voices. Furthermore, the experimental design exploited the English intonation system to ensure that the target segments were produced in the middle of the pitch range. Specifically, all materials for the study were produced on the H* H− intonation pattern (see Beckman & Pierrehumbert, 1986), which is typically used on the items in an open-ended list (Ladd, 1980.) This pattern has a slight rise on the accented syllables, and after the last accent, the high F0 is sustained all the way to the end of the phrase. The F0 does not fall as it would in a neutral declarative pattern; neither does it rise again at the very end, as it would in the typical yes/no question pattern. The pattern in question is illustrated in Figure 3–1. Given these controls, it was possible to access the location and extent of glottalization related to segmental allophony.

The object of the study was to evaluate the reliability and extent of glottalization for vowel-initial words and for voiceless stops in coda position as a function of segmental and prosodic context. The study extends a previous study of voice source allophony under variations in phrasal prosody which was reported in Pierrehumbert and Talkin (1992). Pierrehumbert and Talkin investigated production of /h/ and of glottal stop in a V-V hiatus, but not glottalization of voiceless stops. Their experimental design used male speakers and contrastively low intonation in order to apply an acoustic index of vocal fold abduction for /h/

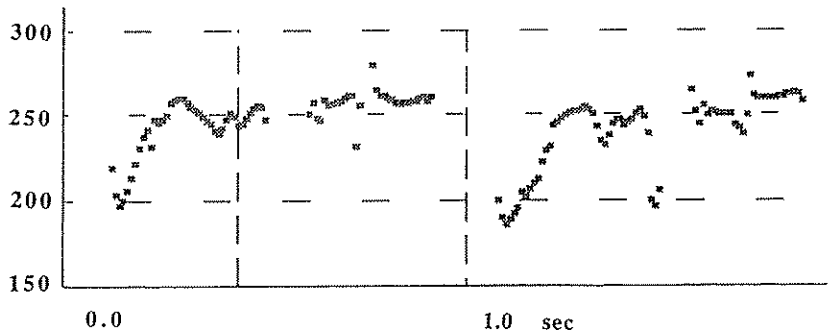


Figure 3–1. F0 contour illustrating the H* H− intonation pattern. Shown are the last two phrases (. . . *Cramer* wattmeters, *mini-* wattmeters) from a three-element list. The context places focus on the modifier; see below. The vertical cursors delimit the region corresponding to the deaccented word "wattmeters" in the first of the two phrases shown; the F0 is high over this region even though it lacks prominence in the phrase.

which did not require inverse filtering. (Inverse filtering is problematic for breathy sounds.) This design proved to be suboptimal for evaluating the glottal stops, and the results on glottal stops were only suggestive. Nonetheless, the study clearly established that an intonational phrase boundary favors glottalization at a vowel-vowel hiatus for the two speakers studied, even if there is no pause at the phrase boundary. For both subjects, expressions such as examples 2 and 3 (in which the second phrase begins with a weak vowel) exhibited at least some glottalization at the boundary in more 90% of the tokens which were continuously voiced through the boundary, without any pause.

(2) Now Emma, _abundance is on the way, variety is on the way, and complexity is also on the way.
(3) Emma, _augmentation is a possibility, reduction is a possibility, and complete removal is also a possibility.

In contrast, in similar segmental contexts which did not involve a phrase boundary, such as examples 4 and 5, glottalization was found in less than 45% of the tokens.

(4) We anticipated the lentil abundance, but not the lima _abundance.
(5) Is this a plasma _augmentation device?

(As careful choice of intonation pattern kept the F0 low in both cases, this difference is not an artifact of the intonation.) A parallel but less extreme difference was found when the second phrase began with a stressed syllable. These results in themselves demonstrate that phrasal prosody is relevant to the segmental allophony.

The study also suggested, but did not firmly establish, that glottalization is favored when the main stress of the phrase falls at the possible glottalization site. In the present study, the consequences of main phrasal stress are brought out more clearly by examining the speech of female speakers toward the middle of the pitch range, instead of that of male speakers toward the bottom of the pitch range.

Both Pierrehumbert and Talkin (1992) and the present paper are embedded in a larger literature dealing with what stress really is from a phonetic and from a cognitive point of view. As is well known, unstressed or weakly stressed syllables are reduced compared to maximally stressed ones; but what does this reduction actually consist of? Lindblom (1963) focused debate on this issue by proposing that reduction effects in weak prosodic positions are mediated by duration. That is, durations for weakly stressed elements are shorter, and shortening has the result that gestures interact more and fail to reach their targets. Harris (1978) provided evidence against this claim by showing that durational differences due to stress and those due to overall speech rate have different consequences for the formant structure; in addition, electromyography (EMG) data were presented indicating that stressed syllables are produced with greater force of articulation. Subsequent literature has supported Harris's point of view, demonstrating that stress involves a more forceful or accurate articulation; see review in de Jong (1990). However, understanding of this issue is still far from complete. The present study contributes on two fronts, by clarifying what properties of the articulation are exaggerated under stress, and by clarifying where in the speech stream this exaggeration occurs.

2. METHODS AND RESULTS

The aim of this study was to explore prosodic effects on the glottalization at a V-V hiatus and at voiceless stops in coda position. However, the extent of such glottalization also depends on the phonetic character of the immediately adjacent segments. This dependence is not well documented in the literature. It was therefore necessary to first conduct a baseline study of the effects of segmental context to permit the design of properly controlled materials for the main study. Here, we first summarize the baseline experiment and then present the main experiment and its results.

2.1. Baseline Study

The baseline study and its implications are reported in more detail elsewhere (Pierrehumbert, 1994). The study, which used the same speakers as the present study, used compounding to vary the segmental context of /t/ and /p/ in coda position. That is, the allophony in compounds such as "hat rack", "campfire", "chopsticks", and "nightgown" was examined to determine which segmental contexts favor glottalization. Glottalization was evaluated objectively but informally by examining the speech waveform and wide-band spectrograms. The primary index of glottalization was irregular periodicity; in cases of very slight irregularity, the spectral balance and the auditory percept were also taken into account. The study established that glottalization is far less prevalent than one might have expected from the literature, at least in the speech of the individuals studied. Many voiceless stops which are frequently transcribed as glottalized displayed no irregularity whatsoever. The only place where glottalization was rather reliably present was for /t/ preceding a sonorant consonant (e.g., in compounds such as "hat_rack", "sprint_ races", "print_maker"). /p/ was sporadically glottalized before /m/ only (e.g., "swamp_maple"). Contrary to the report of Higginbottom (1964), neither /t/ nor /p/ was found to be glottalized before a fricative (e.g., in "Flint_stones" or "camp_fire"). This is not surprising, because voiceless fricatives involve vocal fold abduction and, thus, coarticulation across the word boundary would tend to suppress adduction on the voiceless stop. It should be noted, however, that Higginbottom studied a different dialect which may well glottalize far more aggressively. Nonetheless, these findings suggest that weakly manifested coda stops may often be mistranscribed, perhaps because of a confusion between "glottal stop" and "nothing" in the minds of speakers of English; this result underscores yet again the importance of examining objective measurements of speech.

2.2. Design of Main Experiment

Materials for the main experiment were developed in the light of the baseline experiment, eliminating all but one context in which glottalization is rarely found. As in the baseline experiment, compounding was used to control the syllable structure and the segmental context. The compounds examined exemplified the following phonological situations:

Post-vocalic /t/ preceding a sonorant consonant:

(6) watt_meter, scout_master, knot_weed, foot_wear, boot_leggers, date_loaf

Post-nasal /t/ preceding a sonorant consonant:

(7) font_manager, grant_money, bent_wood, front_wards, tent_lily, mint_leaves

Post-vocalic /p/ preceding a sonorant consonant:

(8) map_maker, tape_measure, stop_watch, lap_wings, sip_lid, step_ladder

V-V hiatus across a word boundary:

(9) Calgary _airfield, silvery _eyeshadow, canary _oilskins, hairy _earworms, savory _eggrolls, cherry _armrests

V-V hiatus across a word boundary in combination with coda /t/ preceding a sonorant consonant:

(10) watery _oat_meal, Gary _Art _Museum, Margery _At_wood, library _art_work, hickory _eight_leaf, cheery _out_look

Note that to ensure the comparability of different sets of forms, all sets involving a voiceless stop before a sonorant consonant contain two forms with /m/, two with /w/, and two with /l/. In addition, the V-V hiatus in every case arises from the presence of the unstressed syllable /ri/ just before the target word. This control was adopted because the treatment of the hiatus appears to depend on the relative quality of the two vowels, and further investigation of this issue would certainly be warranted.

Target sequences were embedded in short discourse segments which strongly motivated a sustained high intonation pattern in the region of interest. As shown by Ladd (1980), the pattern is felicitously used when the speaker wishes to suggest that he or she is providing merely a few examples from a list that could go on much longer. Before the experiment began, subjects were drilled in the desired pattern using examples that did not involve the glottal allophony under investigation here. Example 11 illustrates two discourse segments actually used in the experiment; the target region is again indicated with an underscore.

(11a) It's amazing how bad the breakfasts were there. They had greasy sausages, watery _oat_meal, stale rolls . . .

(11b) We offer the latest looks from head to toe. We have designer knits, high-fashion foot_wear, creative accessories . . .

In the two examples above, the target region is under the main stress in its phrase, because the compound noun is novel, even contrastive, in context. For each discourse segment like those in example 11, a matching segment was constructed that strongly motivated a shift of phrasal stress to the modifier. Given the principles of English intonation, this means that the compound noun was under much reduced phrasal stress. However, because of the choice of intonation pattern, the F0 was at essentially the same level as when the stress was later in the phrase, in fact, often higher.

(12 a) It's amazing how many different ways they found to make bad oatmeal. They had unsalted oatmeal, watery _oat_meal, burnt oatmeal . . .

(12 b) This store has every kind of footwear for the whole family. It has athletic footwear, high-fashion foot_wear, everyday_footwear . . .

As can be seen by comparing examples 11 and 12, the target region in every case is in the middle phrase of a three-item list. The words comprising the middle phrase are identical in the nuclear stress condition (11) and in the post-nuclear condition (12). The only thing that differs is the larger context, which serves to guide the stress placement. Because each of the 30 target phrases was produced in two discourse contexts, a total of 60 discourse segments was constructed. These were arranged in a blocked randomization pattern that counterbalanced for all factors in the experiment. Related discourse segments were separated by at least 14 other segments.

High-quality digital recordings were made in an IAC booth using a Shure SM81 microphone and an Ariel Proport sampling at 16,000 samples per second. The resulting speech signals were examined using the Entropics xwaves+ software package.

2.3. Analysis of the Speech and Results

The primary mechanism for production of glottal stops or glottalized stops is thought to be hyper-adduction of the vocal folds. This results in an irregular excitation pattern involving extra-long pitch periods. A clear

example is shown in Figure 3–2 (a glottalized V-V hiatus taken from the sequence "Calgary airfield").

In addition to creating irregularity of the voice source, vocal fold adduction is also expected to give rise to a sharp discontinuity in the source waveform at the point when the folds come together. A sharp discontinuity translates into a source spectrum with relatively greater energy in the higher harmonics that excites the formants, and relatively less energy in the fundamental. (See Sundberg & Gauffin, 1979.) Because adduction of the vocal folds reduces losses to the subglottal system, glottalization should furthermore give rise to unusually narrow formant bandwidths. However, Pierrehumbert and Talkin (1992) found that irregularity in the excitation was the most reliable hallmark of what is conventionally transcribed as glottalization. The expected spectral correlates were often absent even in regions with a very disturbed source pattern. As a result, irregularity was the primary objective feature examined in the present study.

However, spectral features were examined in conjunction with periodicity, and the materials in the present study did show some of the expected spectral correlates of glottalization. For full glottal stops, narrow F1 bandwidth was quite evident; compare in Figure 3–3 the F1 ringing for the glottal stop in the first panel to the more abrupt deterioration of F1 for an oral stop in the second panel. There was some evidence for enhancement of the high frequency region in glottalized regions.

In interpreting source irregularity, it is important to bear in mind that, vocal fold adduction is not the only source of irregularity. Irregularity can also arise from low airflow and a number of other causes. The consequences of low airflow are of particular relevance here because, when coda /t/ does have an oral closure, this closure will obviously inhibit the airflow.

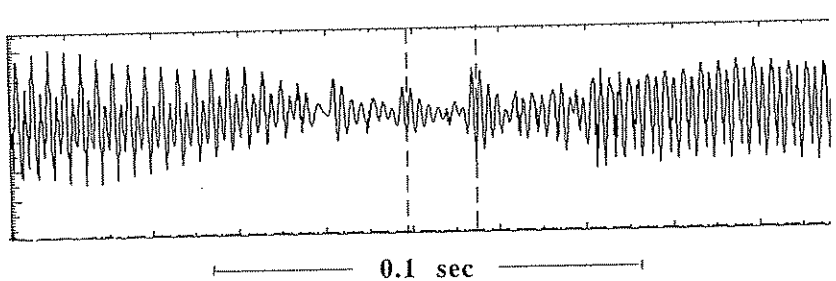
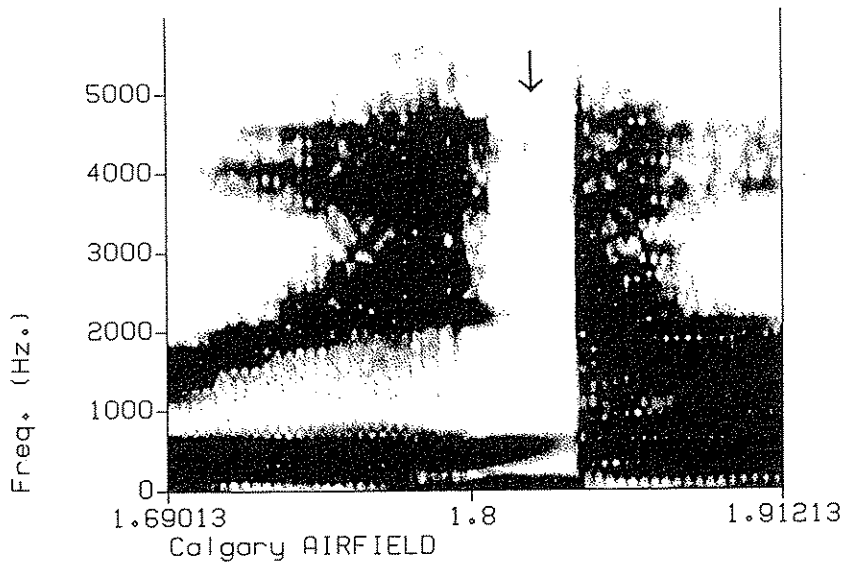
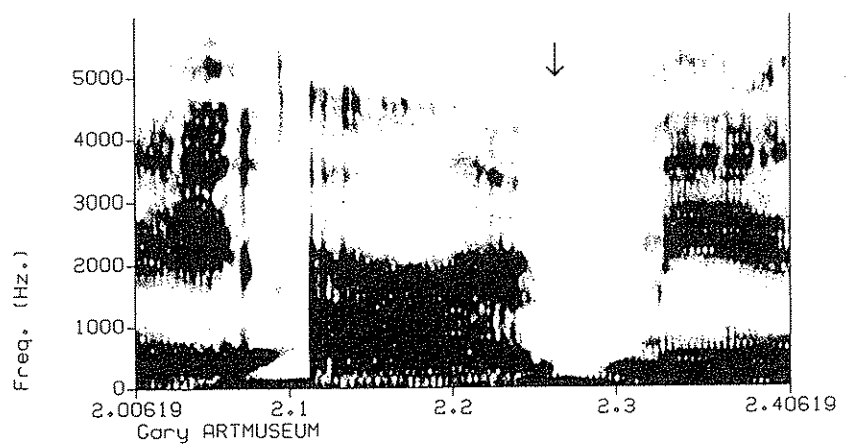


Figure 3–2. Glottalized V-V hiatus in the phrase "Calgary airfield".



A



B

Figure 3–3. A. Spectrogram displaying narrow F1 bandwidth for a full glottal stop in "Calgary _airfield". **B.** Spectrogram displaying more abrupt termination of F1 for the oral stop for "Gary Art_museum".

As a result, any mathematical index of irregularity would not have the same physiological interpretation across all situations in the study.

In view of this difficulty, no attempt was made to develop an all-purpose measure of irregularity. Instead, pairwise comparisons between forms differing only in their prosody were made using the capability of the Entropics xwaves+ software for displaying parameters from several speech files simultaneously. When the segmental context is kept identical in this fashion, it is reasonably clear which speech signal is more glottalized, as shown in Figures 3–4 and 3–5. Comparisons can then be tabulated and subjected to a sign test (a conservative nonparametric test for statistical significance). Comparisons that could not be made with reasonable assurance were recorded as equal; this included particular cases in which one /t/ in a pair appeared to have an oral closure (as indicated by formant transitions and F1 bandwidth), while the other did not.

An extremely robust and significant contrast was found for glottalization breaking up a V-V hiatus. For subject AB, 9 of 12 V-initial words were more glottalized under nuclear stress, 2 were equally glottalized in both stress conditions, and only 1 was more glottalized in post-nuclear position. Nine of 11 full glottal stops observed were found on items

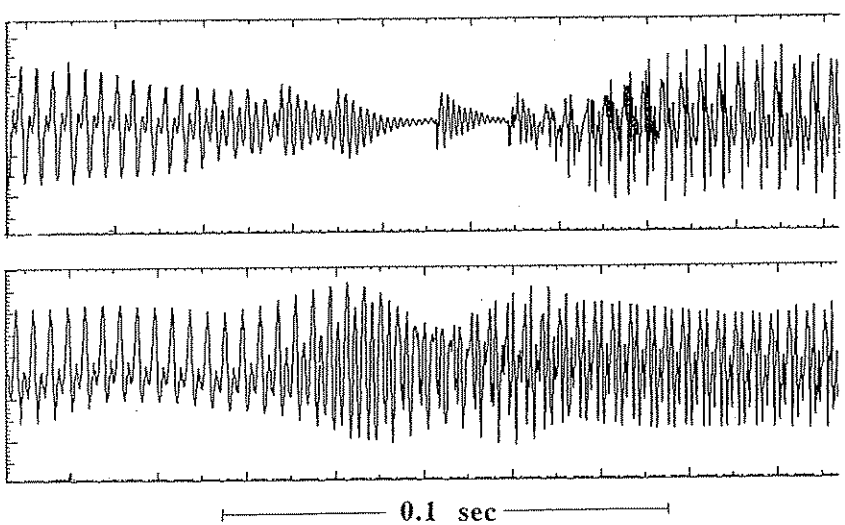


Figure 3–4. Illustration of a pairwise comparison for a V-V hiatus. The upper panel shows the target region of the phrase "silvery _eyeshadow", with main stress on the word "eyeshadow". The lower panel shows the comparable region in "*silvery* _eyeshadow". In this pair, the first shows (much) more glottalization.

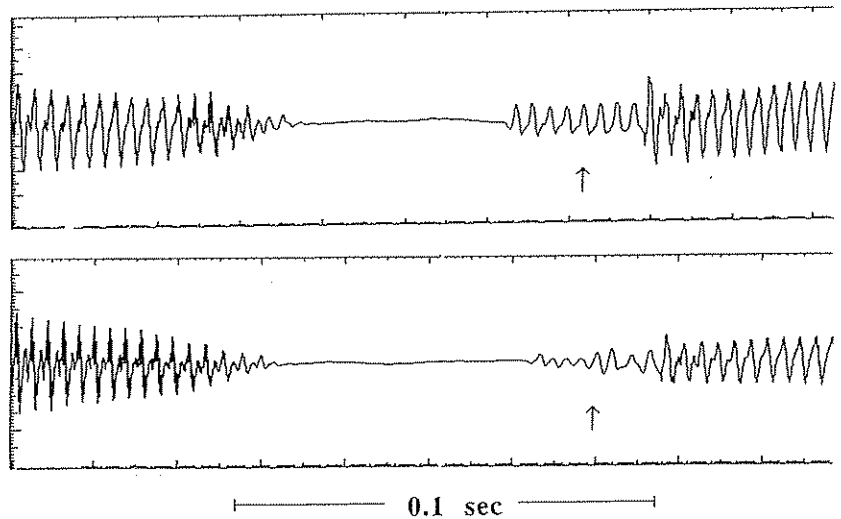


Figure 3–5. Illustration of pairwise comparisons for /t/. The upper panel shows the target region in the phrase "*hickory* eight_leaf"; the bottom panel shows the comparable region in "hickory eight_leaf." Although the difference is very slight, the second was rated as more glottalized because of greater irregularity in the vicinity of the release.

bearing nuclear stress. For subject ST, 10 vowel-initial words were more glottalized under nuclear stress, 1 was equally glottalized in both positions, and 1 was more glottalized in post-nuclear position. All seven of the full glottal stops observed were found on items with nuclear stress. For both subjects, results are statistically significant by a sign test. A sign test is very conservative in this application because it takes no account of the extent of the differences observed, which for most pairs was quite large.

For glottalization of coda stops, the outcome was quite different. Of 24 pairs presenting a candidate for glottalization, AB actually glottalized the stop in at least one member of 20 pairs (all /t/s and two pairs involving /p/ before /m/, replicating the finding of the baseline study.) Of the 20 pairs in which glottalization did occur, 6 showed more glottalization in nuclear position, 6 showed more in post-nuclear position, and in 8 the degree of glottalization was indistinguishable. (There is 100% probability that this difference is due to chance.) For subject ST, the same 20 of the 24 relevant pairs showed some evidence of glottalization. Nine showed more glottalization in nuclear position, 7 showed more in post-nuclear position, and 4

appeared equally glottalized. The small difference between nuclear and post-nuclear position is far from statistically significant.

A vivid illustration of the contrast between the two target positions is found in the behavior of the word "oatmeal" for speaker ST. When "oatmeal" appeared in nuclear position, it began with a full glottal stop, as shown in the upper panel of Figure 3–6. In post-nuclear position, this glottal stop was completely gone, with the result that the beginning of the syllable "oat" was far less glottalized than the end, as shown in the lower panel of Figure 3–6.

3. DISCUSSION

If stress results in exaggerated articulation, what exactly gets exaggerated ? This question may be divided into two parts—where in the speech stream the exaggeration occurs and what aspects of the speech get exaggerated. We will consider each of these questions in turn.

In this study, we have found that phrasal stress strengthens glottalization found word-initially in a V-V hiatus. It does not affect glottalization occurring as a secondary articulation on a syllable-final voiceless stop. These findings may be compared to findings for other phonetic properties under similar manipulations. Pierrehumbert and Talkin (1992) find that phrasal stress strengthened /h/ articulation for initial /h/s in words with initial stress ("hogfarmer", "hawkweed"); similar results for Swedish /h/ are reported in Gobl (1988). In the same words, the vowel is less breathy and has higher amplitude under nuclear stress. In contrast, when /h/ began a weak syllable in a word (as in "tomahawk" and "Omaha"), the magnitude of the /h/ gesture was less affected by whether the word was under nuclear stress or not, though there was some effect. Thus, there appears to be a parallel between the coda /t/s in the present study and the weak onset /h/s in the Pierrehumbert and Talkin study. De Jong (1990) used X-ray microbeam data to investigate jaw, tongue, and lip position for oral stops and for vowels. Contrastive emphasis was used to vary the position of the phrasal stress, as exemplified in example 13.

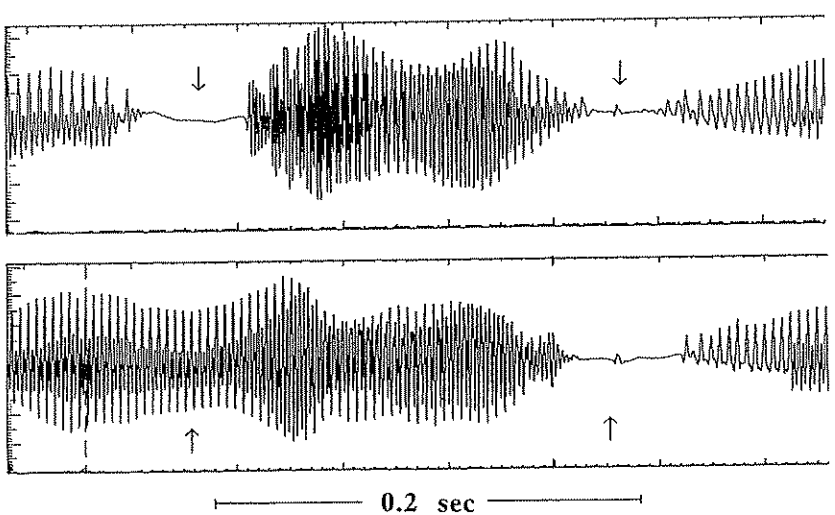


Figure 3–6. The upper panel shows the waveform in the region of the syllable "oat" in "watery _oat_meal". The lower panel shows the comparable region from "*watery* _oat_meal." (Note that the time scale is contracted in comparison to Figures 3–4 and 3–5.)

(13a) I said, "PUT the toad on the table"
(13b) I said, "Put the TOAD on the table"
(13c) I said, "Put the toad ON the table"

Clear evidence was found for hyperarticulation of the labial and coronal features in word-initial position under nuclear stress. Stressed vowels had an increased opening of the front part of the oral cavity; this was accomplished in different ways by different subjects. There is also evidence for more complete articulation of the /t/ in "put" and the /n/ in "on" under nuclear stress.

In summary, for words with initial stress, phrasal stress unequivocally results in exaggerated articulation of the initial consonants and of the lexically stressed vowels. However, the picture for other consonants is less clear. De Jong (1990) found an effect on the oral articulation of the final consonants in "put" and "on"; in contrast, both the present study and Pierrehumbert and Talkin (1992) find only a small effect on the extent of the laryngeal gesture for the /h/ in words like "Omaha", and here we find no effect for the glottalization of /t/ in words like "bent_wood". One possibility is that the sensitivity of measurements in the laryngeal studies was marginal, and that all phonemes in weak positions are affected to some extent by phrasal stress. A second possibility builds on the observation that de Jong's coda consonants were final in the domain of the focus, whereas those examined here and in Pierrehumbert and Talkin (1992) were not. For instance, in example 13a,

"put" is emphasized, but the region of emphasis does not run into the next word ("on"). In the /h/ experiment, the weaker /h/s were medial in a word bearing phrasal stress, and in the present experiment, the entire compound was under phrasal stress, so that the target /t/ was medial in the domain of the phrasal stress. That is, it is possible that de Jong's materials had the coda consonants in a prejunctural position and that the lengthening induced by the boundary reduced the amount of gestural overlap between the vowel and the consonant. On this matter, see also the discussion of focus in Pierrehumbert and Beckman (1988) and de Jong, Beckman and Edwards (1993), and the results on preboundary lengthening in Beckman, Edwards, and Fletcher (1992).

If the second interpretation is the correct one, we may then generalize that the exaggerated articulation induced by phrasal stress is greatest at the onset of the domain carrying the stress and at the head position. Prosodically subordinate positions are much less affected; however phrasal stress may induce phrasing differences which are manifested in longer, but not larger, articulatory gestures.

Let us turn now to the issue of what aspects of the speech are exaggerated. De Jong (1990) compares two approaches, the "sonority profile" model and the "hyperarticulation" model. The concept of the sonority profile was proposed in Silverman and Pierrehumbert (1990) and further developed in Edwards and Beckman (1988) and Beckman, Edwards, and Fletcher (1992). The claim is that the canonical pattern of sonority for the syllable (namely, closed or nonsonorant at the beginning, then open or sonorant, then less sonorant again) is more fully achieved under prosodic prominence, with vowels becoming more open and consonants becoming more closed. As de Jong (1991) notes, this approach fails to account for the increases in accuracy found under stress which have no relation to the overall degree of closure. For example, it does not explain effects of stress such as blocking coarticulation between dentals and alveolars and exaggeration of lip-rounding in stressed rounded vowels. As a result of such observations, de Jong rejects the sonority profile model in favor of a hyperarticulation model. Specifically, he claims that phrasal stress results in hyperarticulation of lexically contrastive features and of allophonic features that enhance contrasts. An example of this last type is lengthening of vowels before voiced obstruents, which is found by Summers (1987) and de Jong (1991) to be exaggerated under phrasal stress. This is not lexically contrastive because vowel length alone does not distinguish between words of English; it is contrast-enhancing because it creates a perceptually available difference between syllables ending in voiced obstruents and those ending in voiceless obstruents.

However, the hyperarticulation model runs into difficulties with the present data on laryngeal allophony and with the data reported in

Pierrehumbert and Talkin (1992). Consider first the pattern found for /h/ by Pierrehumbert and Talkin, as displayed in Figure 3–7. This figure plots the ratio of the energy in F0 to the energy in the first harmonic above the fundamental (or the "second harmonic" in the standard terminology of the Acoustical Society of America.) Because the experimental design ensured that F0 was less than one third of the first formant, the harmonic ratio (or HR) is a good index of the vocal fold abduction which characterizes /h/. HR is relatively high in a very strong (but still voiced) /h/; it is relative low in a well articulated vowel. The figure plots the HR in the /h/ against HR in the following vowel. In each panel, one dotted diagonal line is the line $y = x$ (or where the data points would lie if the /h/ and the vowel were identical, exhibiting no contrast in vocal fold abduction). The other line is a reference perpendicular to $y = x$; the perpendicular distance provides an index of the degree of contrast between the /h/ and the vowel, or the extent of the abduction gesture for the /h/ in comparison to that for the vowel. The plotting figure "A" is used for tokens with nuclear accent, or nuclear stress, on the target /h/; the plotting figure "D" represents tokens in which the target /h/ was in post-nuclear position, or deaccented.

In the top panel, we see that the chief effect of accent on a word-initial /h/ is to increase the contrast between the /h/ and the vowel; data points labeled A are further from the $y = x$ line than data points labeled D. This pattern provides a clear example of hyperarticulation. In contrast, the lower panel shows the results for the /h/s in "Omaha" and "tomahawk". In this panel, A's are somewhat further than Ds from the diagonal, as noted above; however, there is a second and far more substantial effect, namely that A's are displaced towards the lower left of the figure (or parallel to the $y = x$ line) in comparison to the D's. This means that when words like "Omaha" and "tomahawk" carry nuclear stress, both the /h/ and the following vowel are more vocalic than when the words do not carry nuclear stress. This effect, which Pierrehumbert and Talkin term a "background" effect, cannot be explained by hyperarticulation. It would fall under the "sonority profile" model on the assumption that the canonical sonority profile for a focus domain has a sharp rise in sonority at the beginning followed by a slow decline in sonority, spread out over more than one syllable. Further support for this model is found in the host of lenition rules which apply to intervocalic consonants in falling stress configurations in English, shifting them toward more sonorant allophones. Specifically /t/ and /d/ are converted to a sonorant (a flap) in this environment, other voiced stops lenite in the direction of voiced fricatives, and voiced fricatives are reduced towards glides. Although de Jong's model could capture the fact that contrasts are reduced by such allophony rules, it fails to capture the fact the con-

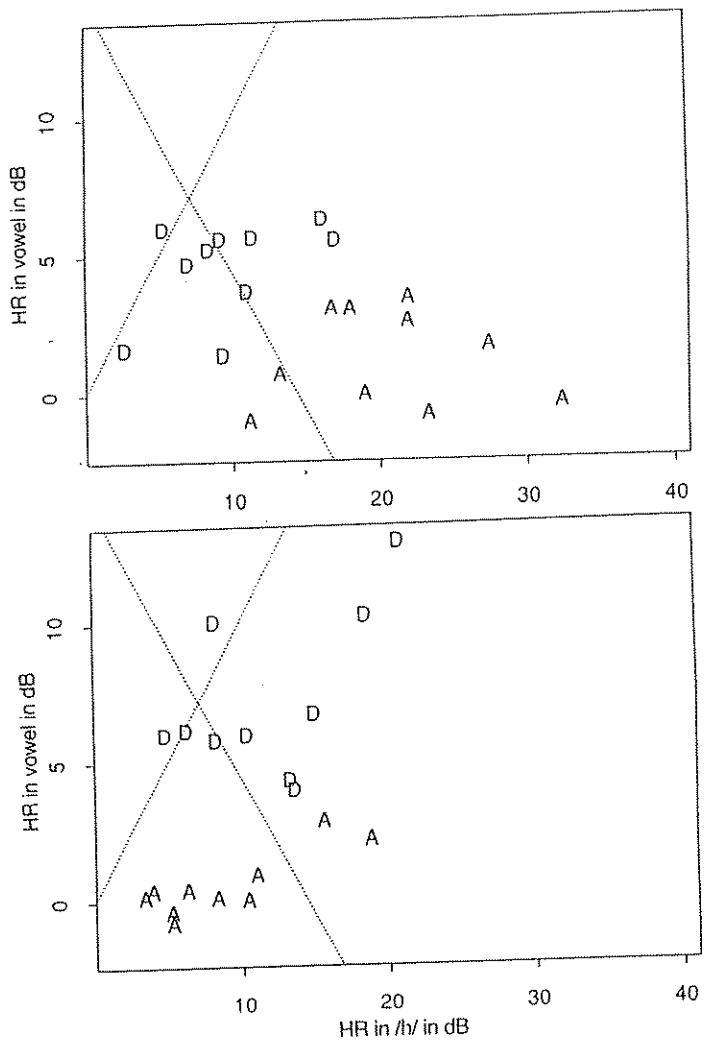


Figure 3–7. Data on the HR for /h/ against the HR for the following vowel, in varying prosodic contexts. Reproduced from Pierrehumbert and Talkin (1992). One dotted diagonal line is the line $y = x$, and the other is a reference perpendicular to $y = x$. "A" represents tokens with main phrasal stress on the word containing the "h"; "D" represents tokens in which the target /h/ was in a deaccented word, that is, in a word following the main stress. The top panel displays data for /h/s that begin a word which has word stress on the first syllable (as in "hog-farmer"). The lower panel displays data for /h/s that begin a noninitial weak syllable (as in "Omaha").

trasts are reduced specifically by pushing all target consonants into more sonorant realizations.

Consider now the word-initial glottal stop found in situations of V-V hiatus. Because glottal stops are not lexically contrastive in English, this glottal stop is inserted by an allophony rule. However, this allophony rule is not reasonably characterized as contrast-enhancing. That is, it does not enhance the contrast between consonant-initial words and vowel-initial words; quite the opposite, it makes vowel-initial words resemble consonant-initial words by supplying them with a consonant. Even on the (unrealistic) assumption that glottal stops are always perfectly distinguishable from all other consonants, the rule is at best neutral with respect to the support it provides for lexical contrasts. Nonetheless, these glottal stops behave just like "real" consonants in terms of being more strongly articulated under phrasal stress. This finding can be accommodated by building on the traditional view of this allophony rule, namely that it supplies a default consonant to provide vowel-initial words with more canonical syllable structure. In short, the observed exaggeration of the glottal gesture can be viewed as exaggeration of phonetic manifestations of canonical syllable structure in accordance with the "sonority profile" model.

A third difficulty with de Jong's formulation is that it does not shed light on the case in which exaggerating the manifestations of one contrastive feature interferes with exaggerating the manifestations of another. This is particularly an issue when we consider the interaction of laryngeal and supralaryngeal gestures. For example, /b/ is both distinctively voiced in English (in contrast with /p/) and distinctively labial (contrasting with /d/ and /g/). The degree of articulation of [labial] is reflected, as de Jong's article demonstrates, by the extent of compression of the lips. Lightly compressed lips might possibly permit some leakage, which would tend to sustain the voicing; indeed, lenition of /b/ in the direction of a bilabial fricative is a well-known strategy for maintaining voicing. Tight compression would preclude leakage and would thus tend to interfere with this strategy for maintaining voicing. Tight compression might also interfere with passive expansion of the oral cavity (another means of sustaining voicing) by increasing the stiffness of the cavity walls. In fact, as Keating (1984) demonstrates, for most English-speaking subjects voicing of /b/ is reduced under phrasal stress; the typical pronunciation of /b/ as a stressed word onset is a voiceless unaspirated bilabial stop. That is, expression of [+voice] has been sacrificed to expression of (+labial) in this position. Keating's data show that for most speakers, /b/ is more fully voiced in a weak position, intervocalically in the onset of a weak syllable. However, a few of Keating's speakers

enhanced (+voice) under stress and according to Beckman (personal communication) Appalachian dialects typically enhance [+voice] under stress by making a kind of implosive. So, the strategy adopted when features conflict is dialect- or idiolect-specific.

This same dilemma is found in a more complicated form in the case of glottalization of /t/. Arguably, glottalization of voiceless stops is an expression of the feature [−voice]. That is, voicing can be suppressed either by abduction or by adduction of the vocal folds. Abduction might be disfavored as a strategy in the context of nasals because (as discussed in Ohala, 1983 and in Ohala & Ohala, 1993) breathy voicing introduces a side-branch into the vocal tract whose spectral consequences are highly confusable with spectral consequences of the side branch created by nasalization. That is, the result of aspiration would not be perceptually salient as it would be confusable with spectral properties that were found independently in the context. This line of reasoning might also extend to a lateral context. This would leave adduction as the mechanism for [−voice] in the vicinity of nasals and laterals, with extension perhaps by analogy to other sonorant consonants. However, adduction results in long and irregular vocal pulses, with the result that formants are not reliably excited during their most rapid transitions. That is, glottalization interferes with one of the main cues to place for a post-vocalic stop, namely the formant transitions going into the closure. In fact, for glottalized /t/, it is often impossible to determine from the acoustic record whether a coronal closure has actually occurred. Some dialects systematically omit the coronal closure. When it is necessary to speak very clearly to resolve an ambiguity, most speakers will use an aspirated released allophone of /t/ instead of the glottalized one. Altogether, these observations suggest that for /t/, expression of the feature [coronal] is being sacrificed to expression of [−voice] in this situation.

4. CONCLUSION

This study establishes that the production of word-initial glottal stops is more extreme under phrasal stress. This provides further support for the claim that phrasal stress has broad effects on allophony and it means that word-initial glottal stops pattern with lexically contrastive consonants in this position. Furthermore, stress is found to have a much greater effect on the extent of word-initial glottalization than on glottalization of voiceless stops in coda position.

Comparison of these results to other results in the literature indicates that neither the "sonority profile" model nor the "hyperarticulation" model can explain all phonetic consequences of stress. Instead, the results suggest that hyperarticulation under phrasal stress interacts with canonical manifestations of prosody to determine the actual outcomes. Thus, the results of this study confirm the point of view in Pierrehumbert and Talkin (1992) in which consequences of stress are decomposed into effects on the magnitude of contrast and effects on how contrasts are located within the relevant articulatory range.

One of the main questions for the sonority profile model is how consonants in prosodically weak positions behave in comparison to those in stronger positions. That is, if the opening gesture of the syllable is exaggerated by making the onset less sonorous and the nucleus more sonorous, how does the closing gesture fare? In fact, is the syllable really the only relevant domain or also some larger unit? The present results, in combination with those reported elsewhere, suggest that the consonants in weak positions are much less affected than either the onsets of stressed syllables or stressed vowels. These weak consonants include both the coda stops studied here and the onsets of weakly stressed syllables studied in Pierrehumbert and Talkin (1992) and Keating (1984). Hence the canonical shape of units larger than the syllable is shown to be relevant.

ACKNOWLEDGMENT

This work was supported by NSF Grant No. BNS-9022484. I am grateful to Mary Beckman for reviewing the initial draft of this paper.

DISCUSSION AFTER PRESENTATION

R. McGowan: There may be no contradiction in compressing the lips more during bilabial closure and voicing. Voicing can continue through extra movement of the larynx, velum, or expansion of the pharyngeal walls.

J. Pierrehumbert: My suggestion was that, everything else being equal, tight lip compression tends to interfere with voicing. Certainly, it does not absolutely preclude voicing; speakers can overcome the interference by adopting more complex articulatory strategies. Keating's data show that most subjects do give up voicing of /b/ when it is under stress, but a few do maintain the voicing. This minority is probably using some of the strategies you mention.

K. N. Stevens: Couldn't the initial glottal stop enhance the following vowel? It gives well defined formants from the beginning of the vowel. This is an alternative account to the sonority view.

J.P.: This is an extremely good point which is supported by vocal source modeling and deserves to be pursued. I have just a few reservations about this to suggest. First, in Pierrehumbert and Talkin (1992), a study involving male subjects at the lower end of their pitch range, the expected effects of glottalization on spectral balance were not found, and we were forced to abandon spectral balance as a measure. Second, in examples in the present study which had a full glottal stop at a V-V hiatus, the major effect on voice source was observed on the unstressed syllable /ri/ preceding the stop, rather than on the stressed syllable itself. Third, the glottal stop does have perceptually significant hallmarks of a voiceless stop (e.g., a stop gap and a spectral edge upon release) and these make it potentially confusable with other voiceless stops, especially in noise. The trade off between this factor and the enhancement of the vowel quality needs to be explored.

O. Fujimura: What about /h/? Is that also a vowel?

K.N.S.: (h) also has some of the properties of a vowel, but it would not greatly affect the adjacent vowel.

O.F.: Glottalization is a rather complex set of phenomena, and from a physiological point of view, it can be produced in many different ways. The difference in the production mechanism may or may not be related to the observed variety of functions. One thing I would like to mention is that Masayuki Sawashima and myself, a number of years ago, did some simple fiberoptic experiments on final /t/ in American English, and we found the false vocal folds to be actively adducted in that situation. Some of the contexts for the final /t/ in your experiment may involve that particular gesture also. It's reported in the University of Tokyo RILP Annual Bulletin (Fujimura & Sawashima, 1971; Fujimura, 1990).

J.P.: One thing I really would appreciate are suggestions from you all about how to parameterize this better in view of the complexity of the situation. So if anyone has any relevant studies, I would like to know about them.

REFERENCES

Anderson, M. (1986). *Pitch determination of speech signals*. Unpublished master's thesis, Department of Electrical Engineering and Computer Science, Massachusettes Institute of Technology, Cambridge.

Beckman, M., Edwards, J., & Fletcher, J. (1992) Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II* (pp. 68–86). Cambridge, U.K.: Cambridge University Press.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 15–70.

de Jong, K. (1990). *The supra-glottal articulation of prominence in English: Linguistic stress as localized hyperarticulation*. Manuscript submitted for publication.

de Jong, K. (1991). An articulatory study of consonant-induced vowel duration changes in English. *Phonetica 48*, 1-17.

de Jong, K., Beckman, M., & Edwards, J. (1993). The interplay between prosody and coarticulation. *Language and Speech, 36*, 197–212.

Edwards, J., & Beckman, M. (1988). Articulatory timing and the prosodic interpretation of syllable duration. *Phonetica 45*, 156–174.

Fujimura, O. (1990). Methods and goals of speech production research. *Language and Speech, 33*, 195–258.

Fujimura, O. & Sawashima, M. (1971). Consonant sequences and laryngeal control. *Annual Bulletin, Research Institute of Logopetics and Phoniatrics, University of Tokyo, 5*, 1–6.

Gobl, C. (1988). Voice source dynamics in connected speech. *KTH Speech Transmission Laboratory, Quarterly Progress and Status Report, 1*, 123–159.

Harris, K. (1978). Vowel duration and its underlying physiological mechanisms. *Language and Speech, 21*, 354–361.

Higginbottom, E. (1964). Glottal reinforcement in English. *Transactions of the Philological Society*, London, 129–142.

Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language 60*, 286–319.

Ladd, D. R. (1980). *The structure of intonational meaning*. Bloomington: University of Indiana Press.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America, 35*, 1773–1781.

Ohala, J. (1983). The phonological ends justifies any means. In S. Hattori & K. Inoue (Eds.), *Proceedings of the 13th International Congress of Linguists* (pp. 232–243). Tokyo: Sanseido Shoten.

Ohala, J., & Ohala, M. (1993). The phonetics of nasal phonology. In M. K. Huffman & R. A. Krakow (Eds.), *Phonetics and Phonology 5: Nasals, Nasalization, and the Velum* (pp. 225–250). New York: Academic Press.

Pierrehumbert, J. (1994). Knowledge of variation. Invited talk at CLS 30. Paper to appear in the Papers from the 30th Regional Meetingof the Chicago Linguistics Society, University of Chicago, Chicago.

Pierrehumbert, J., & Beckman, M. (1988). Japanese tone structure. *Linguistic inquiry monograph 15*. Cambridge, MA: MIT Press.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Doherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II* (pp. 90–117). Cambridge, U.K.: Cambridge University Press.

Silverman, K., & Pierrehumbert, J. (1990). The timing of prenuclear high accents in English. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I* (pp. 72–106). Cambridge, U.K.: Cambridge University Press.

Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America, 82*, 847–863.

Sundberg, J., & Gauffin, J. (1979). Waveform and spectrum of the glottal voice source. In B. Lindblom & S. Ohman (Eds.), *Frontiers of speech communication research*. London: Academic Press.

CHAPTER

4

A Phonation-Type Synthesizer for Use in the Field

Peter Ladefoged

1. INTRODUCTION

This paper reports a somewhat eccentric endeavor: to design, construct, and test a method of synthesizing differences in voice quality in vowels that uses only acoustic parameters and works in real time on a portable computer. The eccentricity of this notion may not be immediately apparent, as the utility of a system of this kind is clear. It would obviously be useful if there were a tool that enabled clinicians or linguists to go off to hospitals and schools, or remote areas such as the Kalahari Desert, and study voice qualities of patients or native speakers of little-known languages, noting how they compare with a standard set of voice qualities as produced on a computer. The eccentricity comes when we require this to be done using strictly acoustic parameters in real time on a portable computer.

We are so used to describing speech in terms of the acoustic theory of speech production (Fant, 1960), that we tend to forget that this is a description that does not rely on acoustic notions alone, as it makes reference to physiological entities. It separates the properties of the vocal fold source from the properties of the vocal tract transfer function. Virtually all speech synthesizers do this in some way, in that they consider the source function not as an impulse but as having particular