A preliminary perceptual study of the vowels of Montana Salish: The method of adjustment as a fieldwork technique

Keith Johnson*
Peter Ladefoged

Montana Salish, a Southern Interior Salish language, has a 5 vowel system /i/, /e/, /ɑ/, /o/, /u/ which is an elaboration of the historical Salishan 4-vowel system /i/, /ɑ/, /u/, /ə/ (Thompson, 1979). In addition, the language, like other Salishan languages, has labialized back consonants. In this paper we present a preliminary study of the vowel system which was aimed at identifying the internal perceptual structure of the vowel categories and in particular the influence of back consonant labialization and vowel stress on the perceptual/phonetic properties of /o/. We begin with a brief description of the method of adjustment perception task and the rationale for its use in field phonetics.

The most common method of instrumental phonetic fieldwork is acoustic analysis. Speakers are recorded in the field and the recorded utterances are later analyzed in the laboratory. Two problems arise with this method. First, speaker-specific characteristics (arising from anatomy and personal style) are confounded with linguistic properties. This is especially problematic when utterances produced by only a few speakers are taken to represent a language. Second, speech is a product of abstract linguistic/phonetic goals and concrete articulatory constraints on vocal movements, and it is difficult to separate the goals from the constraints.

The strategy normally taken to overcome these difficulties with acoustic analysis is to record many utterances produced by a large number of speakers. A critical mass of acoustic data which includes variation in speakers and in speaking styles makes it possible to pull apart speaker and situational variability to some extent, but this sort of critical mass is often difficult to acquire in a field situation. Therefore we have been experimenting with a new approach to the problem of separating the effects of speaker and language, and the effects of linguistic targets and the contingencies of speech production (Johnson, Flemming & Wright, 1993). Our strategy is to have listeners adjust some parameters of a speech synthesizer until a synthetic speech sound matches a particular sound in the language. The results of this method of adjustment perceptual task are speaker-independent in the sense that each listener hears the same synthetic 'speaker'. It is still necessary to use a critical mass of listeners, but all listeners, even those with very different vocal tract shapes, react in a similar way to the synthetic voice, so only a small number of listeners is required. In addition, previous research (Johnson et al., 1993) suggests that when listeners perform the task they tend to 'undo' reduction processes found in normal speech production.

* Department of Linguistics, Ohio State University, 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210-1298.

This last claim needs a little elaboration. Johnson et al. (1993) found that American English listeners in a method of adjustment perception task chose vowel sounds which were different from the vowels they produced in normal speech. The vowels chosen in the task had more extreme formant values expanding the perceptual vowel space relative to the measured acoustic vowel space. Johnson et al. also showed that this expanded perceptual vowel space corresponded to vowels produced in hyperarticulated or 'clear' speech. Several experimental manipulations (using groups of listeners with different linguistic training, giving different instructions to the listeners, and varying the number of redundant cues available in the stimuli) had no impact on the results, indicating that the vowel space expansion observed in the task was not an artefact. This finding was interpreted as evidence that the method of adjustment task taps abstract (though physically definable) linguistic goals. Of course, data acquired in the method of adjustment task do not (and cannot) replace the information obtained by acoustic analysis. For instance, the perception task reveals nothing about dynamic aspects of vowels, patterns of vowel reduction in normal speech, or differences between men's and women's speech. Still, this technique does provide a useful control over two confounding variables in traditional acoustic analysis, and thus supplements such an analysis.

One purpose of the present study was to explore the applicability of the method of adjustment in a field situation. Our previous research involved a rather limited group of listeners (university students) so we had no way of knowing how listeners less oriented toward test-taking would perform the task. They might not find the listening task sufficiently interesting to be able to complete it; or there might be other cultural factors which made it difficult for them to understand what to do. Consequently it was not clear whether the method of adjustment would be useful in field phonetics.

Beyond these methodological questions we were interested in the effects of contextual variability on perceptual vowel categories. Miller & Volaitis (1989) found that goodness rating functions, a measure analogous to the method of adjustment, for voice onset time (VOT) are sensitive to rate of speech. This evidence that the internal structure of a perceptual category can be modified by phonetic context challenges the conclusion reached by Johnson et al. (1993) who did not study whether varying the phonetic context of the vowel to be matched had any effect. In this study of Montana Salish, we explored the effect of phonetic context in the method of adjustment task by investigating three variants of /o/; stressed, unstressed and contextually labialized.

Method

The study was conducted in the context of a brief two day visit to the Cultural Center on the Flathead Reservation in Montana, thanks to the assistance of the Director of the Flathead Culture Committee, Clarence Woodcock. The principal purpose of the visit was to record an extensive word list illustrating the phonological contrasts of Montana Salish. The present experiment was conducted during intervals in this process with the help of the three main language consultants: Harriet Whitworth, Felicite ("Jim") McDonald, and Dorothy Felsman, to whom we are very grateful.

Listeners

Three female native speakers of Montana Salish served as listeners in the study. All three were also fluent speakers of English. Their ages ranged from mid 50s to mid 70s.

Materials

Words illustrating the five vowels of Montana Salish in stressed syllables were selected (Table I). Table I lists two additional words; in one the vowel /o/ appears after a labialized consonant and in the other /o/ is in an unstressed syllable. These words were included in the list to investigate the role of stress and contextual labialization on perceptual vowel quality. Table I shows the order in which the words were presented. If lack of familiarity with the task has any impact on the listener's performance we expect this familiarity effect to be largest for the first words in the list.

Table I

Words illustrating the vowels of Montana Salish. The first column lists the vowel which the word was used to illustrate, the second column lists the words in Montana Salish orthography, and the third column lists the English glosses. Words are arranged in the order in which they were presented in the perceptual task.

/é/	č'upélsi	'lonesome'
/ɨ/	píləlš	'a stagger, stumble'
/ʷo/	mxʷól'	'cradle-board'
/á/	tám	'it's not, wrong'
/ó/	pólplqən	'thimbleberry'
/ú/	púlsəm	'he killed something'
/o/	olín	'belly'

The computer-produced synthetic vowel sounds used in the experiment have been described in more detail in an earlier paper (Johnson et al., 1993). A set of 298 separate vowel sounds were synthesized using a software formant synthesizer (Klatt & Klatt, 1990). The first vowel formant (F1) varied from 250 Hz to 900 Hz in 15 equal steps, while the second vowel formant (F2) varied from 800 Hz to 2799 Hz in 22 equal steps. The step-sizes were calculated on a psychoacoustic scale (the Bark scale), so they represent equal auditory distances rather than equal acoustic distances. F1 was constrained to be at least 300 Hz below F2; consequently not all possible combinations of F1 and F2 values (e.g. F1 = 900 Hz and F2 = 800 Hz) were used. A regression formula (Nearey, 1989) was used to calculate the value of F3. F4 was fixed at 3500 Hz unless F3 was above 3200 Hz in which case F4 was 300 Hz higher than F3. The bandwidths of F1-F3 were also calculated by regression formulas derived from values used to synthesize vowels in English. The fundamental frequency was 120 Hz with a slight fall at the end of the 250 ms long stimulus.

Procedure

The experiment was run on-line using a Macintosh Powerbook 140. The visual display presented to the listeners was composed of a grid of (15 X 22) 330 square buttons

on a Hypercard card and the instruction, for example 'Find the é vowel of č'upélsi', was printed at the bottom of the screen. The words were presented in the order shown in Table I. Clicking on one of the squares caused the corresponding stimulus from the F1/F2 vowel array to be played to headphones or a separate amplifier and speaker. The sampling rate was 22 kHz with 8 bit quantization.

The listener's task was to search through the stimuli to find the computer-synthesized vowel that most closely matched the sound of the specified vowel in that word. Even though the listeners had no previous experience with a computer, all three were quickly able to perform the task. We suggested a strategy to the listeners; namely that they start each trial by sampling different quadrants of the grid and then continue to work within a quadrant to find the preferred stimulus. HW (the first listener) had some difficulty in controlling the cursor with a trackball, so the other two listeners used a mouse. Also, the stimuli were presented over a loud speaker to the first two listeners, but we considered that room noise and the experimenter's reactions to the stimuli might impact performance so listener DF heard the stimuli over headphones. Each listener performed the task once for each of the words listed in Table I, except HW, who had to skip the last word.

Results

Before discussing the results it is important to note that this is a preliminary study. As the method section indicated, we collected a very scant amount of data; only one repetition of the perceptual task for the seven example words, with only three speakers of the language. Therefore, this section is merely a record of how we went about trying to use a new perceptual technique, what lessons we learned, and what types of linguistic phonetic questions we think can be answered using this method. The results suggest that a more extensive study would indeed shed valuable light on the phonetic structure of Montana Salish.

The results are shown in Figure 1. The ellipses show 65% bivariate confidence intervals for the mean formant values. Between-listener variance was greatest for the vowels /i/ and /ɑ/, and was also rather large for /o/ if the different allophones are pooled together. On the other hand, the listeners chose very similar values for the vowels /u/ and /e/. Note that, with the exception of /i/, the range of formant values observed in this experiment was comparable to or smaller than the measured male formant values reported by Flemming, Ladefoged, & Thomason (1994).

If the words illustrating /i/ and /ɑ/ had been presented first in the list we could have attributed the between-listener variation to lack of familiarity with the task. However, the first word in the list illustrated /e/ (see Table I), and, as the figure shows, the listeners chose very similar formant values for this vowel. It is therefore not obvious that the large variability seen for /i/ and /ɑ/ is the result of uncertainty on the first few trials in the experiment. It may also be important that the listeners did not generally disagree with each other; their choices for the vowels /e/ and /u/ were similar even though listening conditions (loudspeaker vs. headphones) and response methods (trackball vs. mouse) differed. This

suggests that the method of adjustment task may be used in field situations to provide a concise speaker-independent view of a language's vowel space. However, before we can confidently rely on the method of adjustment it is necessary to perform the same experiment with several other speakers and to have each speaker repeat the task several times for each word in order to get some idea of the range of between- and within-listener variation one is likely to encounter.

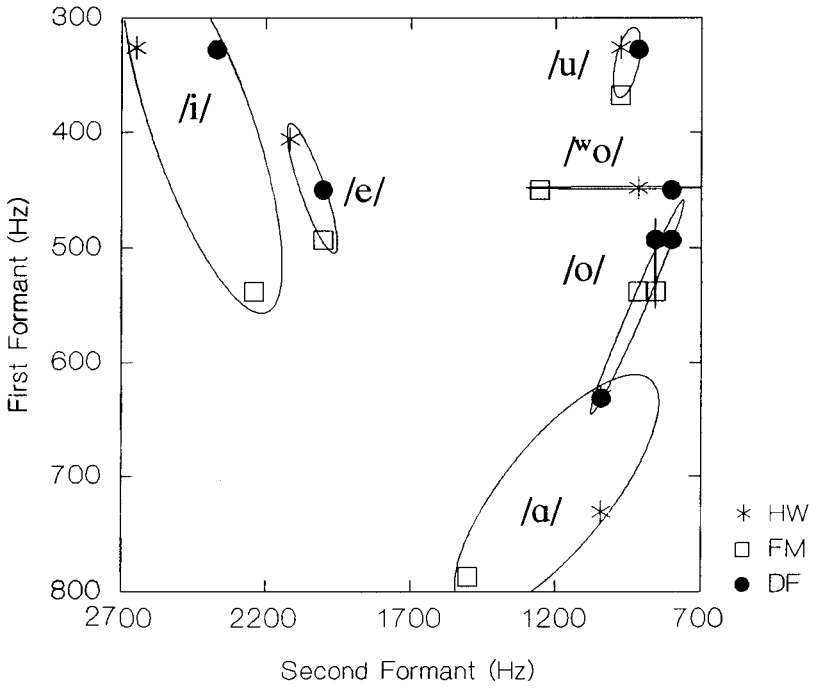

Figure 1. Results of method of adjustment study of Montana Salish vowels. Formant values of the synthetic vowels chosen for the words listed in Table I are plotted by listener.

We were also interested in the three variants of /o/. The stressed and unstressed variants are not separately labeled in Figure 1 because they overlap with each other. The stressed /o/'s are connected by the diagonal ellipse and the unstressed /o/s are connected by the vertical ellipse/line. This overlap of stressed and unstressed versions of /o/ indicates that even though vowel quality is (to our ears) much reduced in unstressed syllables in normal productions, listeners tended to equate the stressed and unstressed versions of /o/ in the perceptual task. Interestingly however, HW and FM seem also to have maintained a distinction between two other variants of /o/; one with a preceding labialized consonant and one without. Although DF chose a lower F1 for /ʷo/, her choices for the three versions of /o/ had very similar formant values. So, HW and FM seem to have kept /ʷo/

and /o/ separate while DF did not. This difference between the speakers is correlated with a difference in how they listened to the synthetic vowel sounds. HW and FM listened to the sounds over a loudspeaker, in a room that was not always quiet, while DF listened to the sounds over headphones and in a less busy room. Obviously, further research is needed before any conclusion can be drawn.

Discussion

From this very brief study we learned several methodological lessons. (1) The method of adjustment is a feasible tool for phonetic fieldwork. The listeners seemed to find the task at least moderately interesting (as compared with the complete and total boredom induced by more common speech perception tasks), and the results were provocative and interpretable. (2) To get usable data it will be necessary to collect several repetitions of each vowel from each speaker. Some estimate of within-speaker variation is needed, and we need to see if variability decreases as experience with the perceptual task increases. (3) A hearing screening is needed. Our listeners were above 50 years of age and may have had some hearing loss. (4) Headphones are preferable to loudspeakers.

The results of the present experiment on the different allophones of /o/ are provocative. Miller & Volaitis (1989) reported an effect of speaking rate on the internal structure of phonetic categories. Johnson & Armstrong (in preparation) also found an effect of F0 on the vowel space obtained in the method of adjustment. The effect of contextual labialization found in this study is, however, of a different sort. Speaking rate and F0 (or speaker identity) persist over long stretches of speech, while coarticulatory effects have much smaller domains (see Johnson, 1991 for an example of the differential perceptual effects of these two types of phonetic 'context'). Therefore, it is perhaps not surprising that the internal structure of a perceptual category might be affected by global nonlinguistic properties such as rate and speaker. However, if the method of adjustment taps a somewhat abstract level of representation, as suggested by Johnson et al. (1993), we have to consider why contextual labialization in the visually presented word could have an effect on the method of adjustment result (see also Volaitis & Miller, 1992).

The overlap of stressed and unstressed /o/ in the listeners' responses suggests that these two versions of the phoneme are considered to be the "same" and may have the same phonetic target (Johnson, et al., 1993). The pattern shown by listeners HW and FM, who kept /o/ distinct /ʷo/ leads to the conclusion that either the variant of /o/ in labialized context has become phonemicized in Montana Salish or that this experimental technique taps knowledge at a subphonemic level. The pattern shown by listener DF suggests that more data is needed before any clear answer will emerge. It is possible that some listeners match sounds at a somewhat abstract level of representation, whereas others pay more attention to phonetic details.

References

Flemming, E., Ladefoged, P. & Thomason, S. (1994) Phonetic structures of Montana Salish.

Johnson, K. (1991) Differential effects of speaker and vowel variability on fricative perception. *Language and Speech* **34**, 265-279.

Johnson, K. & Armstrong, A.A. (in preparation) Memory for speaker-specific vowel information in speech.

Johnson, K., Wright, R. & Flemming, E. (1993) The hyperspace effect: phonetic targets are hyperarticulated. *Language* **69**: 505-528.

Klatt, D. & Klatt, L. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* **87**, 820-857.

Miller, J.L. & Voliatis, L.E. (1989) Effects of speaking rate on the perceived internal structure of phonetic categories. *Percept. Psychophys.* **46**, 505-512.

Nearey, T.M. (1989) Static, dynamic & relational properties in vowel perception. *J. Acoust. Soc. Am.* **85**, 2088-2113.

Thompson, L.C. (1979) Salishan and the Northwest. In *The Languages of Native America: Historical and Comparative Assessments*. (L. Campbell & M. Mithun, Eds.) Austin: Univ. Texas Press.

Volaitis, L.E. & Miller, J.L. (1992) Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *J. Acoust. Soc. Am.* **92**, 723-735.