TR - H - 161

The Effects of Talker Variability on the Perception of American English /r/ and /l/ by Japanese Listeners, II: Subject differences, acoustic and temporal correlates of talker effects, and some technical considerations

James S. Magnuson
Reiko A. Yamada

1995. 7. 25

ATR人間情報通信研究所

〒619-02　京都府相楽郡精華町光台2-2　☎ 0774-95-1011

ATR Human Information Processing Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
Telephone: +81-774-95-1011
Facsimile: +81-774-95-1008

The effects of talker variability on the perception of American English /r/ and /l/ by Japanese listeners, II: Subject differences, acoustic and temporal correlates of talker effects, and some technical considerations

James S. Magnuson
Reiko A. Yamada

ATR Human Information Processing Laboratories
2-2 Hikaridai
Seika-cho, Soraku-gun,
Kyoto, 619-02 Japan
email: magnuson@hip.atr.co.jp, yamada@hip.atr.co.jp
Telephone: (+81)7749-5-1084
Fax: (+81)7749-5-1008

A manuscript combining the results reported here and in Technical Report TR-H-110 will be submitted to the Journal of Experimental Psychology: Human Perception and Performance.

ABSTRACT

In an earlier technical report [Magnuson and Yamada, TR-H-110, 1994.12.6], we presented the results of three experiments which investigated the effects of talker variability on the perception of American English (AE) /r/ and /l/ by Japanese adults. In that first report, we found that Japanese subjects set criteria for /r/-/l/ decisions based on the range of cues they experienced in any given block of trials, independently of talker-specific differences in cues to /r/ and /l/. This led to significant differences in how often subjects responded "R" (R-rate) to particular talkers in blocked (single) and mixed (multiple) talker conditions. At that time, we reported that we were unable to find acoustic explanations for the talker effects we observed. We extend and comment on our previous report in three ways. First, we describe a new analysis of subject differences in those experiments. Second, we describe a series of analyses that have revealed some weak correlations between acoustic and temporal measurements and talker effects observed in the first two experiments in the previous technical report. Finally, we describe a minor flaw in the design of the second experiment and recommend a scheme for avoiding similar flaws in future experiments.

This first section is taken largely from Magnuson and Yamada (1994.12.6). Readers familiar with that technical report may wish to skip to the next section.

Differences in talker characteristics are a well-known source of problematic variability in speech perception. Due to differences in age, sex, size, dialect, and other factors, the way different talkers acoustically realize the same linguistic segments may be quite different, and the way they realize different linguistic segments may be quite similar (for example, the range of one talkers' productions of /i/ may overlap with another's productions of /I/ in terms of formant space; see Peterson and Barney, 1952). All the same, while listening to native-language speech, people have little trouble normalizing[1] for talker differences in phonemic cues. Is this also true of non-native phonemes which contrast on dimensions that are not distinctive in the native language?

Logan, Lively and Pisoni (1991) reported talker-specific differences in Japanese listeners' accuracy in an /r/-/l/ identification task, both preceding and following identification training with feedback. The talker-specific patterns appeared in tests of generalization with new stimuli, and persisted even when as many as 45 training sessions were used (Yamada, 1993). This result suggests that individual talkers may give differential emphasis to the multiple cues to /r/ and /l/ (although acoustic correlates to the perceptual differences have not yet been found), and that Japanese listeners have not experienced a sufficient sampling of the range of cues to /r/ and /l/ that occur across different talkers to be able to normalize for this kind of variability in /r/ and /l/ productions. Considering talker differences in cues as adding to the

---

[1] Because of recent descriptions of the "traditional" view of talker normalization as the stripping away of all information not relevant to phonetic decisions (e.g., Palmeri et al., 1993), we should note from the outset that this is not what we mean by normalization. We mean relating (achieving a mapping between) stored mental representations of speech categories to the speech source despite the variability added by the current context (e.g., talker characteristics, room acoustics, etc.). Although some of the mechanisms postulated by Nusbaum and Morin (1992) and others (see Nearey, 1989, for a review) work only on what might be called the dimensions most relevant to phonetic decisions (e.g., formant trajectories), they do not involve stripping from memory any other information carried by the speech signal.

range of cues suggests a connection to previous work by Yamada and Tohkura (1990, 1992) and Underbakke et al. (1988).

While the steady-state onset and frequency transition of F3 is sufficient for native speakers of American English (AE) to distinguish /r/ and /l/ across talkers, /r/ and /l/ also differ systematically in the spectral and temporal characteristics of the first two formants (O'Connor et al., 1957). When Yamada and Tohkura (1990) covaried spectral cues to /r/ and /l/ (F3 and F2), native speakers of AE clearly based their decisions on F3. In contrast, Japanese subjects' responses were influenced by both cues. Underbakke et al. (1988) reported a similar result. They used a trading-relations paradigm to contrast the performance of AE and Japanese subjects given consistent spectral and temporal cues, a single spectral cue, and conflicting spectral and temporal cues. They found that for Japanese listeners, similarity to AE listeners' performance was correlated with accuracy in an /r/-/l/ identification task. Japanese subjects with high initial accuracy were able to make appropriate trade-offs given different sets of acoustic cues, but Japanese subjects with low initial accuracy did not adjust to the different sets of cues. The less skilled Japanese listeners obviously attended to different acoustic parameters than native listeners: they responded more to temporal patterns of F1 than to F3. Underbakke et al. concluded that the correlation of temporal and spectral cues that distinguish /r/ and /l/ for native speakers of English are not artifacts of the human auditory system, but are a function of linguistic experience with language-specific allophonic rules. Yamada and Tohkura (1992) extended these findings by examining the decision processes of Japanese listeners.

Yamada and Tohkura (1992) found that the range of variation in acoustic cues differentiating /r/ and /l/ presented in a particular experimental session affected the labeling performance of Japanese listeners, but not that of native speakers of AE. No matter what portion of a synthesized /r/ - to - /l/ continuum they presented to Japanese listeners, rates of "R"[2] responses were approximately 50%. Given only

[2] We will use phonemic transcriptions (e.g., /r/ and /l/) to denote the intended phonemic category of a speaker or the category to which native speakers would assign a stimulus. We will use quoted upper-case roman characters (e.g., "R" and "L") to denote response categories.

the /r/-half of the continuum, only the /l/-half of the continuum, or the entire range, subjects apparently set labeling criteria such that they responded "R" on approximately half of the trials within a block. Yamada and Tohkura concluded that their Japanese subjects perceived /r/ and /l/ continuously, rather than categorically. Without well-defined categories for /r/ and /l/, Japanese subjects set criteria relative to the range of cues to /r/ and /l/ they heard within a block of trials. This contrasts with the near-categorical criteria native subjects employed with the same stimuli (see Yamada and Tohkura, 1992, for details).

These results suggest that while Japanese subjects with limited experience in an English-speaking environment are able to divide a set of stimuli into "R" and "L" categories, they do not possess robust categorization criteria that they can apply to individual stimuli independently of cue variability. That is, they respond on the basis of relative differences, rather than absolute (categorical) criteria.

Considering the talker-specific accuracy patterns reported by Logan et al. (1991) together with the range effects found by Yamada and Tohkura (1992), we made predictions about the effects of within-session talker variability (Magnuson and Yamada, 1994.12.6). If subjects are unable to normalize for talker differences in non-native speech contrasts, we might predict that to naive Japanese listeners, the relative range of cues to /r/ and /l/ would be increased when stimuli from two or more talkers are presented within a block as compared to when stimuli from only one talker are presented within a block. In analogy to Yamada and Tohkura's (1992) "range effects", we should not observe large changes in the overall rate of "R" response if we manipulate the range of cues to /r/ and /l/ by varying the amount of talker variability. However, if subjects set criteria based on the overall range of cues they hear within a block of trials, those criteria should become less successful as the number of talkers increases, and identification errors will increase.

The experiments reported in our previous technical report were designed to investigate the effects of random talker changes on the perception of /r/ and /l/ by Japanese listeners; specifically, whether Japanese listeners are able to normalize for talker-specific differences in cues to /r/ and /l/, or if added variability due to talker differences

instead influences session-specific criteria, as did the "range" manipulations in Yamada and Tohkura's study (1992).

In Experiment 1, we examined these questions with a paradigm used to study talker normalization processes in native-language stimuli. In native-language speech perception, subjects perform faster or more accurately in word recognition tasks when stimuli are presented blocked by talker than when stimuli from different talkers are mixed in a series of trials (cf. Nusbaum and Morin, 1992; Magnuson, Yamada and Nusbaum, 1994). The additional processing time in the mixed-talker condition has been attributed to a process whereby subjects analyze (normalize for) the characteristics of changing talkers. Nusbaum and Morin (1992) found that subjects responded more quickly in a monitoring task when stimuli were blocked by talker. They also found that talker variability hindered subjects' performance in concurrent memory tasks, which indicates that adapting to talker changes requires attention and memory resources. One measure of whether or not Japanese subjects are able to adjust to changes in /r/-/l/ cues due to increased talker variability is a comparison of performance when stimuli are blocked by talker with performance when stimuli from multiple talkers are mixed in one session. If Japanese subjects do not have well-defined knowledge of the variability that occurs between talkers in the non-native phonemes /r/ and /l/, they may attempt to set session-specific, talker-independent criteria for "R-L" decisions. If cues to /r/ and /l/ in all talkers' productions are not distributed identically on a common set of parameters, this would result in identification errors when stimuli from multiple talkers are mixed in a block of trials.

In Experiment 2, we extended our analysis to an examination of the time course of adaptation to talker changes. Kato and Kakehi (1988; Kato, 1992) reported several observations concerning the adaptation of trained listeners in a transcription task following talker changes. In one experiment, subjects transcribed stimuli in noise. The talker was kept constant for several trials, and then changed. Kato and Kakehi observed sudden decreases in accuracy followed by gradual increases in accuracy (over the course of 3-5 trials) when the talker changed. These results suggest that listeners are able to "tune" their recognition processes to a particular talker. According to this "contextual tuning" theory of talker normalization (Nusbaum and Morin, 1992), subjects analyze a talker's vocal characteristics based on an initial experience

with the talker's speech. As long as the talker does not change, subjects can reference the result of the analysis in working memory more efficiently than they can re-compute vocal characteristics. This allows subjects to redirect more cognitive resources to other attention-demanding experimental tasks. (See Nearey, 1989, and Nusbaum and Morin, 1992, for discussions of mechanisms that may be used to analyze vocal characteristics.)

In Experiment 3, we combined the manipulation of talker variability with a manipulation of the proportion of /r/ and /l/ stimuli presented in a block, in an attempt to replicate Yamada and Tohkura's (1992) "range effects" with natural stimuli. We predicted (and found) that without knowledge of the true proportion or /r/ and /l/ stimuli, Japanese subjects would attempt to divide whatever range of /r/ and /l/ stimuli they hear into two balanced categories.

In this report, we will discuss analyses of subject differences and physical correlates to R-rate effects in Experiment 1 and 2. We will also discuss design issues relevant to Experiment 2. We will not discuss Experiment 3 any further.

EXPERIMENT 1: SUBJECT DIFFERENCES AND PHYSICAL CORRELATES TO R-RATE

Experiment 1 was designed to compare how subjects classify the same stimuli in blocked- (single) and mixed- (multiple) talker conditions; that is, whether stimuli are judged independently of the amount of talker variability within a block of trials. Subjects were asked to identify the initial sound of words that began with /r/ or /l/ produced by five talkers. Subjects performed this task in two talker conditions: a blocked talker condition, in which subjects heard the productions of each talker in five separate blocks; and in a mixed-talker condition, in which the productions of all five talkers were mixed in random order (see Magnuson and Yamada, 1994.12.6, for details of the experimental procedure).

Given previous results (the "range effects" and talker effects discussed above), we predicted that Japanese subjects' rates of "R" response for specific talkers would change significantly between blocked- and mixed-talker conditions. In the blocked-talker condition, we predicted that subjects would set criteria based on the range of only

one talker's productions. In the mixed-talker condition, we predicted that subjects would set a single criterion based on the range of all talkers' productions, with the result that Japanese subjects' classifications of some stimuli would change when judged relative to the entire range.

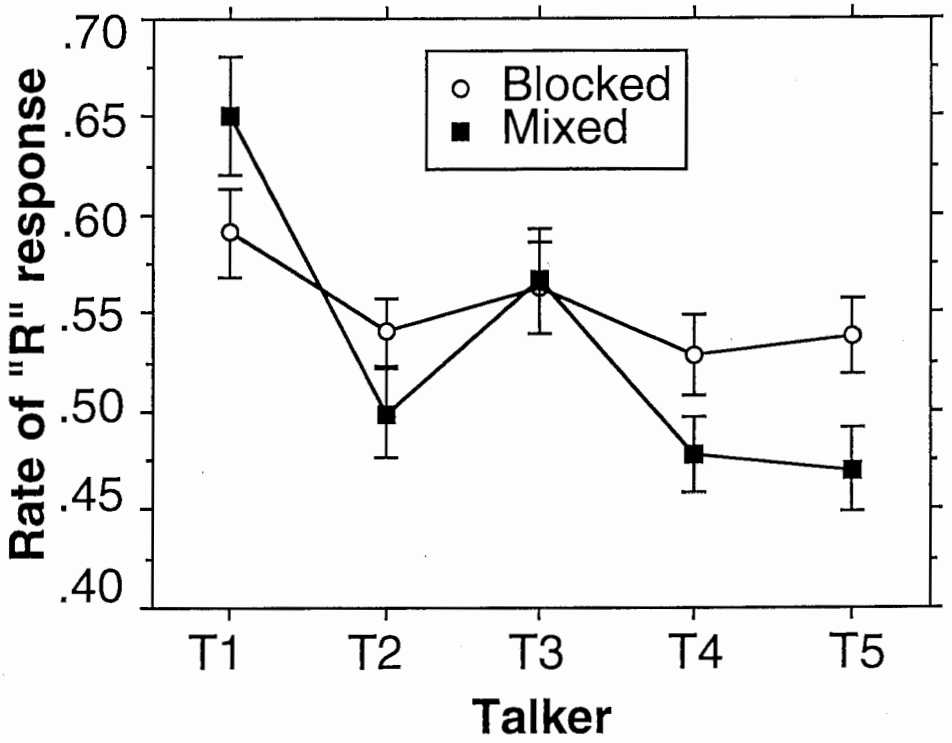


Figure 1: The interaction of talker and talker condition in Experiment 1.

Indeed, we found the predicted effects (see Magnuson and Yamada, 1994.12.6, for detailed analyses). The overall R-rate was close to .50 in every block. However, in the mixed-talker condition, R-rate to particular talkers differed substantially from R-rate in the blocked condition, leading to the significant interaction of talker and talker condition ($F(4,104) = 6.11$, $p < .001$) shown in Figure 1.

In our previous report, based on this result and also detection theory analyses which indicated that the effect was due to response bias, we concluded that subjects were setting criteria based on the range of cues they hear in a block of trials, as was reported by Yamada and Tohkura (1992). Except that in this case, the range was augmented by differences in redundant cues produced by the different talkers. For illustration, consider Figure 2. Suppose three talkers' productions of /r/ and /l/ are distributed along some unknown dimensions, as indicated by

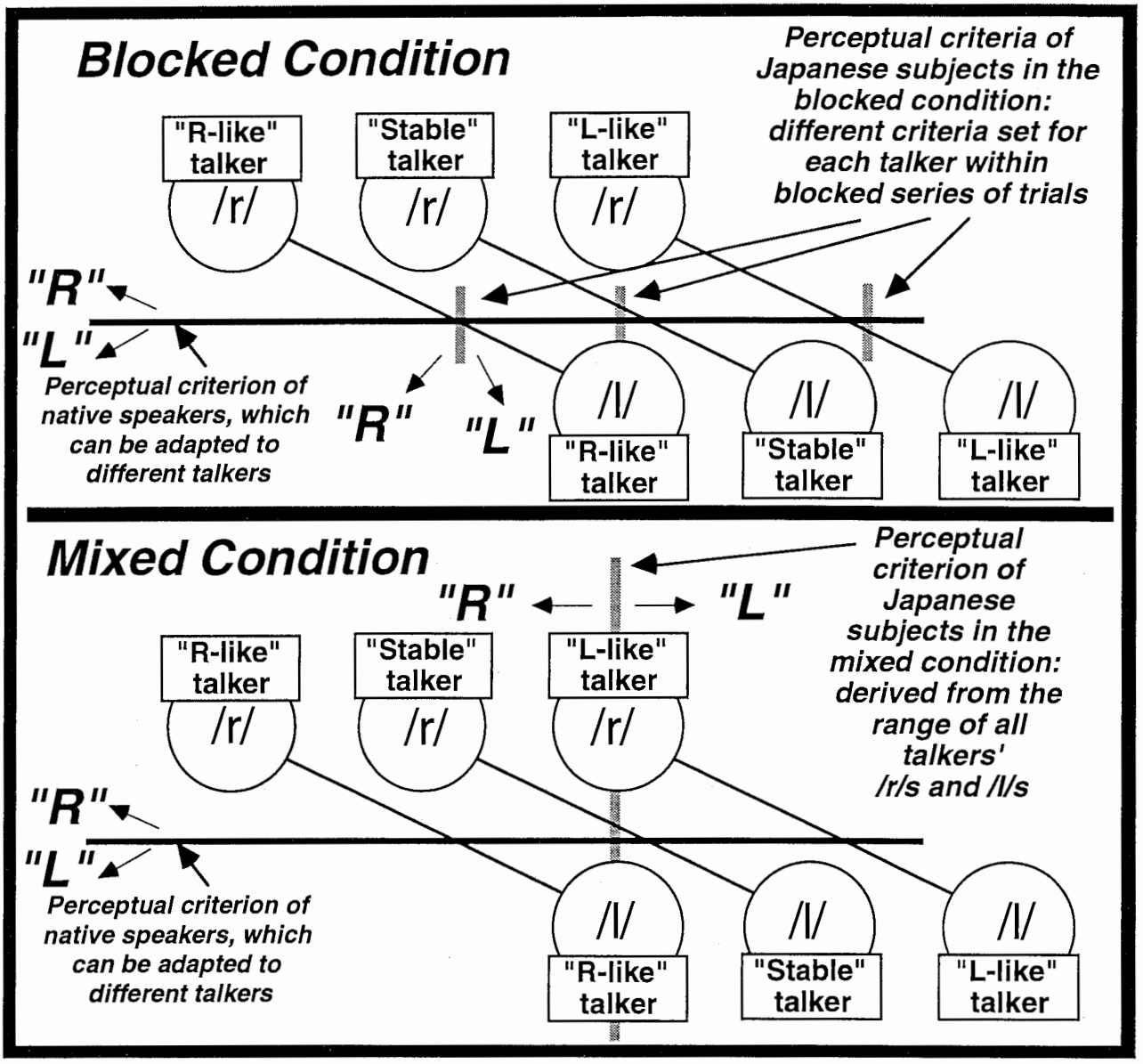


Figure 2: A schematic representation of our hypothesis to explain the interaction of talker and talker condition in R-rate.

the connected circles in Figure 2. Native speakers understand that all of the talkers' productions can be classified according to some common regularity, possibly with some adaptation required for talker differences. This is indicated in both panels of Figure 2 by the heavy horizontal lines, which properly divide all three talkers' distributions of /r/ and /l/. The gray vertical lines represent how Japanese listeners divide the range of cues into two categories. In the blocked condition (upper panel), the criteria chosen seem appropriate to particular talkers, but this is only because the variability in /r/-/l/ cues within the block is due to the particular talker's productions. When the same strategy is applied in the mixed talker condition (lower panel), and a single criterion is set for a block of trials, many more of some talkers' /r/s and /l/s are identified as "R" (the R-like talker), and many more of some talkers' productions are identified as "L" (relative to the blocked-talker condition) -- due to the increased range of possible cues to /r/ and /l/ identity introduced by the talker differences.

Subject differences. Since our last report appeared, we have considered more carefully when and why subjects might use such a strategy. This analysis was inspired by a prediction made by Dr. Yoh'ichi Tohkura. He predicted that subjects could be divided into three groups based on how accurately they are able to identify /r/ and /l/. Two of the groups would not show the talker X talker condition interaction we found in Experiment 1. First, subjects with low accuracy (near 50% overall), would not show such an effect: as they are at chance level, it is unlikely that they have well-enough formed /r/ and /l/ categories to employ such a strategy. Second, subjects with high accuracy would not show such an effect, either, for the opposite reason: to achieve high accuracy, subjects would have to have well-enough formed /r/ and /l/ categories that they would not have to rely on such a strategy. The third group would consist of those subjects whose accuracy would fall somewhere between the "low" and "high" accuracy groups. These subjects would have well-enough formed categories of /r/ and /l/ that they could actively construct session-specific criteria for /r/-/l/ decisions.

We found exactly what Dr. Tohkura predicted. R-rate to each talker is plotted against average accuracy in Figure 3. The significant interaction of talker and talker condition in Experiment 1 was

apparently due to the response patterns of subjects with average accuracy above approximately 53% and less than approximately 86%.

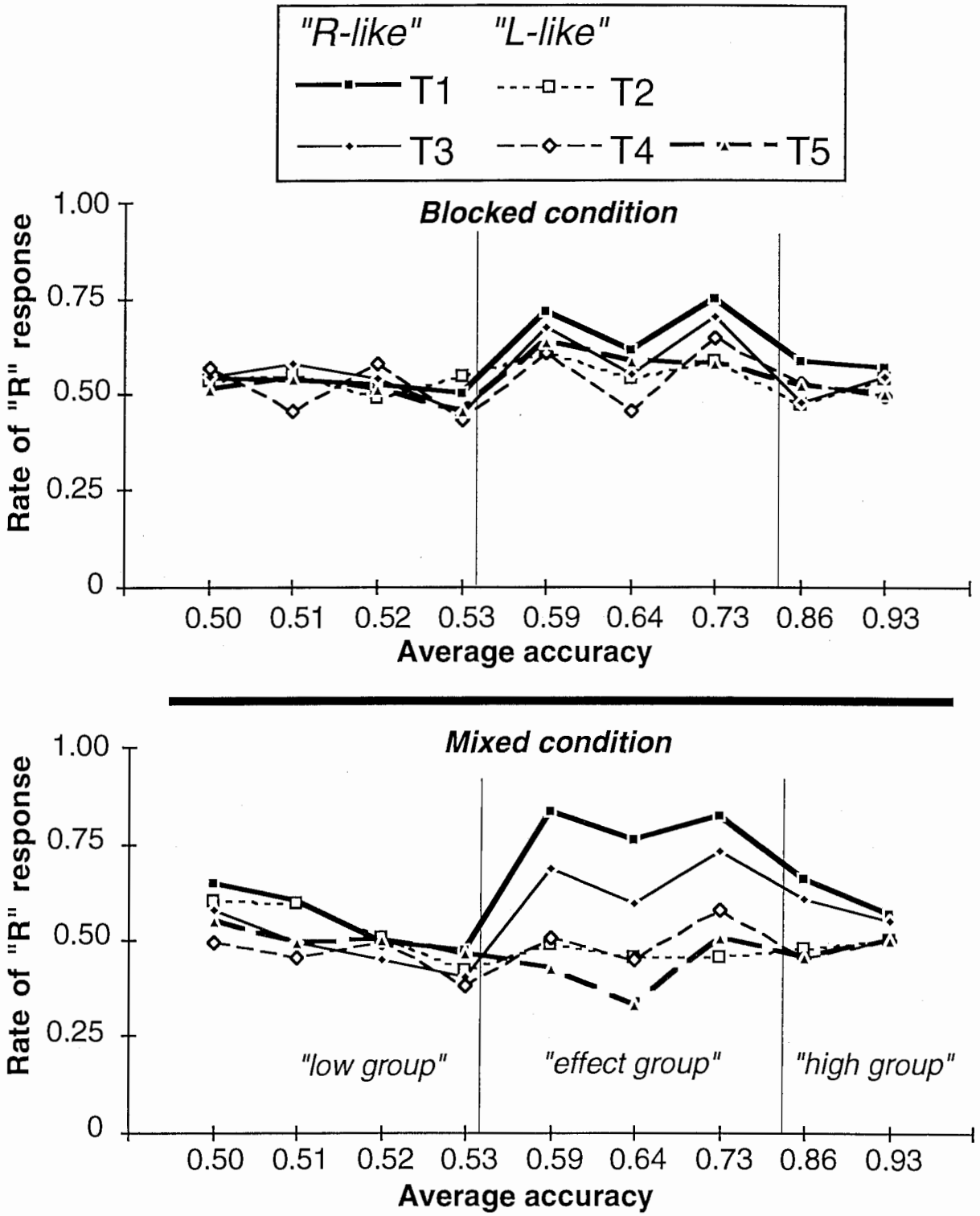


Figure 3: Rate of "R" response to each talker by average accuracy in blocked and mixed talker conditions in Experiment 1. Each point represents the average rate of "R" response for three subjects. Note that the numbers on the X axis are labels only.

Acoustic analyses. We now turn to the question of what cues subjects are using to make "R-L" decisions. We identified a number of possible cues, and examined their ecological validity, and their correlation with subjects' rate of "R" response for each stimulus. The cues examined were initial center-frequencies of F1, F2, F3; duration of the /r/ or /l/ portion of a stimulus; the identity of the following vowel; and the height (high, mid, or low) of the following vowel. F1, F2, and F3 were measured using formant-tracking and smoothing algorithms (written by Seiichi Tenpaku for ATR; the programs used were *ftrack* and *fmt_smooth*). The parameters used are given in Table 1. The LPC order was varied slightly for some talkers' /r/s and /l/s in order to achieve the best results. Outlying values were verified or corrected manually.

Parameter	Value
Sampling frequency	10 kHz
Window length	30 msec
Window type	Hanning
Frame period	5 msec
Pre-emphasis	.98
LPC order	10 to 12

Table 1: Parameters used for formant tracking.

For the stimuli used in Experiment 1, there was one ecologically-valid cue: F3 was completely reliable. The maximum measured value of F3 for /r/ stimuli was 1980 Hz ($\underline{M}$ = 1638 Hz, range = 1031 Hz to 1980 Hz) and the minimum measured value of F3 for /l/ stimuli was 2097 Hz ($\underline{M}$ = 2826 Hz, range = 2097 Hz to 3326 Hz). A simple regression was computed for the relation of consonant identity to F3: $\underline{y} = -.001\underline{x} + 3.102$, $\underline{r}^2 = .853$. The next highest match was with F2: $\underline{y} = -.001\underline{x} + 2.59$, $\underline{r}^2 = .122$. F2 is not a good cue for "R-L" decisions. Although the mean measured values were 1163 Hz for /l/ stimuli and 1040 Hz for /r/ stimuli, the ranges were similar: 753 Hz to 1610 Hz for /l/ stimuli, and 660 Hz to 1475 Hz for /r/ stimuli. Vowel and vowel height were obviously not ecologically-valid cues, since the stimuli were minimal pairs contrasting only in the initial consonant. Nor were F1 and duration reliable cues ($\underline{r}^2 = .0002$ and .001, respectively).

GROUP	Cue	BLOCKED r^2	MIXED r^2	Change in fit: mixed-blocked
ALL	F1	.011	.042	.031
	F2	**.154**	**.184**	.030
	F3	**.559**	**.534**	-.025
	Duration	.020	**.111**	**.091**
	Vowel	.000	.006	.006
	Height	.000	.004	.004
EFFECT	F1	.015	.088	**.073**
	F2	**.190**	**.271**	**.081**
	F3	**.371**	**.362**	-.009
	Duration	.053	**.237**	**.184**
	Vowel	.002	.002	.000
	Height	.001	.001	.000
HIGH	F1	.000	.002	.002
	F2	**.161**	**.199**	.038
	F3	**.787**	**.805**	.018
	Duration	.001	.010	.009
	Vowel	.000	.000	.000
	Height	.000	.000	.000
LOW	F1	.017	.011	-.006
	F2	.001	.006	.005
	F3	.022	.002	-.020
	Duration	.005	.018	.013
	Vowel	.000	.019	.019
	Height	.000	.011	.011

Table 2: Results of simple regressions of cue values to rate of "R" response for each accuracy group in Experiment 1. Values of r^2 greater than .100 are presented in boldface. Change in fit (r^2) is presented in the right-most column, and changes greater than .050 are presented in boldface.

What cues did subjects use? We must note that we cannot determine whether or not subjects ever actually used F3 in Experiment 1: because of the division between /r/ and /l/ stimuli in F3, high

accuracy will always be correlated with F3, even if subjects are relying on another cue we have not identified (but see Yamada and Tohkura, 1990, who were able to analyze sensitivity to F3 by systematically chinning it in synthetic stimuli). That said, the results of simple regressions between rate of "R" response and the six cues are presented by talker condition for all subjects, and by effect group (effect, high, and low) in Table 2. The correlations were between R-rate and the various measures for all stimuli produced by all talkers. When correlations were computed separately for all of the talkers, there were few substantial differences between blocked and mixed conditions or between talkers: in some cases there were large changes in fit to F3 between the blocked and mixed conditions, but this was correlated with accuracy differences. The correlation with duration was never higher than .133 in the individual talker analyses.

Returning to the correlations computed across all stimuli, the two cues with the highest fit to /r/-/l/ identity, F3 and F2, also had the highest fit to subjects' rate of "R" response in both talker conditions. It also appears that subjects in the effect group relied to some degree on the duration of the /r/ or /l/ portion of stimuli in the mixed-talker condition (r^2 = .237), but not in the blocked-talker condition (r^2=.053, change in r^2 = .184), and also relied more heavily on F2 in the mixed condition. The large increase in the fit to duration between the blocked and mixed conditions for the effect group suggests that subjects may well have focused on duration as a cue to apply across talkers in the mixed-talker condition.

Talker	F1	F2	F3	Duration
T1 (R-like)	333.94	1011.08	2035.68	126.76
T2 (L-like)	362.08	1158.38	2323.12	67.08
T3 (stable)	325.36	1055.08	2081.98	93.80
T4 (L-like)	408.26	1114.72	2338.02	75.76
T5 (L-like)	448.06	1169.22	2383.22	68.96

Table 3: Average cue values for each talker.

Average cue values by talker for cues that could vary between stimuli and talkers are shown in Table 3, and from the table it appears that the F1, F2, F3 and duration cues are all correlated with the "R-

like"- or "L-like"-ness of the talkers. However, given that the greatest mixed-blocked change in fit occurred for the duration cue, we will tentatively suggest that duration may be a cue subjects applied across talkers in the mixed-talker condition.

EXPERIMENT 2: SUBJECT DIFFERENCES, PHYSICAL CORRELATES TO R-RATE, AND SOME TECHNICAL CONSIDERATIONS

For Experiment 2, we selected the talker that seemed "R-like" in the mixed condition in Experiment 1 (T1, for whom subjects' rate of "R" response was relatively high in both blocked and mixed conditions), one "stable" talker (T3, for whom subjects' rate of "R" response was not affected by talker condition, but was slightly "R-like", in that his /r/s were easier to identify than his /l/s), and one talker that seemed "L-like" (T2, for whom subjects' rate of "R" response decreased in the mixed condition). T2 was not the most "L-like" talker from Experiment 1; however, we chose these three talkers since all of them were male, in order to avoid the possibility of confounding talker sex effects with effects due to "degree of R-likeness".

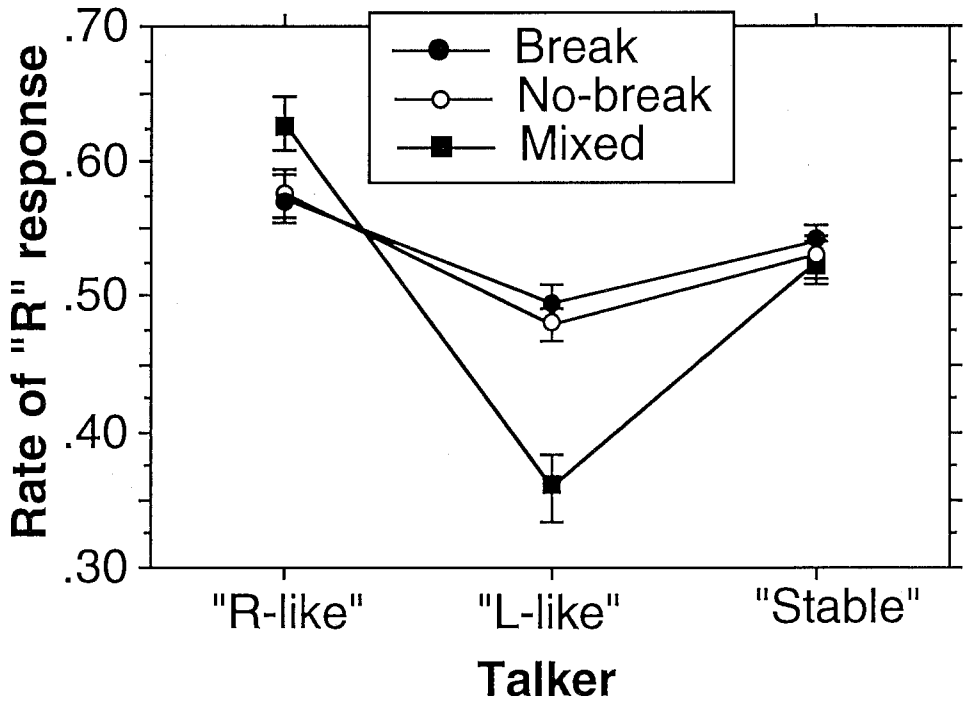


Figure 4: Rate of "R" response in Experiment 2 by talker and condition. Note that there are no differences between the two blocked conditions (break and no-break) for all three talkers; also, the differences between "R-like" and "L-like" talkers are more apparent in the mixed condition. (Bars represent standard error.)

Subjects listened to minimal pairs of words contrasting /r/ and /l/ in initial position from each of the three talkers, in three conditions: mixed (all stimuli from all talkers randomly ordered), blocked-with-breaks (or, the "break" condition, with stimuli from each talker presented sequentially, in random order, with a forced break between talkers in which subjects listened to instrumental music), and blocked with no-breaks (the same as the previous condition, without any sort of pause between talkers changes). In Experiment 1, we also failed to find patterns of gradual tuning after talker changes, such as those found by Kato and Kakehi (1988; see Kakehi, 1992, for a description in English). Since it is possible that it could simply take subjects longer to tune to talkers when listening to non-native speech than the time we gave them in Experiment 1, we increased the number of stimuli from 25 pairs to 50.

As in Experiment 1, we found a significant interaction of talker and talker condition ($F(4,132) = 17.34$, $p < .001$). As can be seen in Figure 4, the interaction of talker condition and talker are due to the rate of "R" response being much lower to the "L-like" talker than to the others in the mixed condition.

Subject differences. In Figure 5, you can see that there were subject differences similar to those found in Experiment 1. The rate of "R" response of the least accurate subjects (average accuracy <~ 51%) did not show the interaction of talker and talker condition apparent in the overall data (Figure 4), as those subjects were operating at chance levels. The most accurate subjects (average accuracy ~ 80%) did not show a strong interaction either, as they were apparently able to adjust to talker-specific differences.

Acoustic analyses. The cues examined for Experiment 1 were examined for the set of stimuli used in Experiment 2. The cues were initial center-frequencies of F1, F2, F3; the duration of the /r/ or /l/ portion of a stimulus; the identity of the following vowel; and the height (high, mid, or low) of the following vowel. F1, F2, and F3 were measured using a formant-tracking algorithm (see Table 1 for parameters). Outliers were verified or corrected manually.

As in Experiment 1, F3 was the only ecologically-valid cue. The maximum measured value of F3 for /r/ stimuli was 1920 Hz ($M = 1586$ Hz, range = 1031 Hz to 1980 Hz) and the minimum measured value of F3 for /l/ stimuli was 2097 Hz ($M = 2687$ Hz, range = 2097 Hz to 3336 Hz).

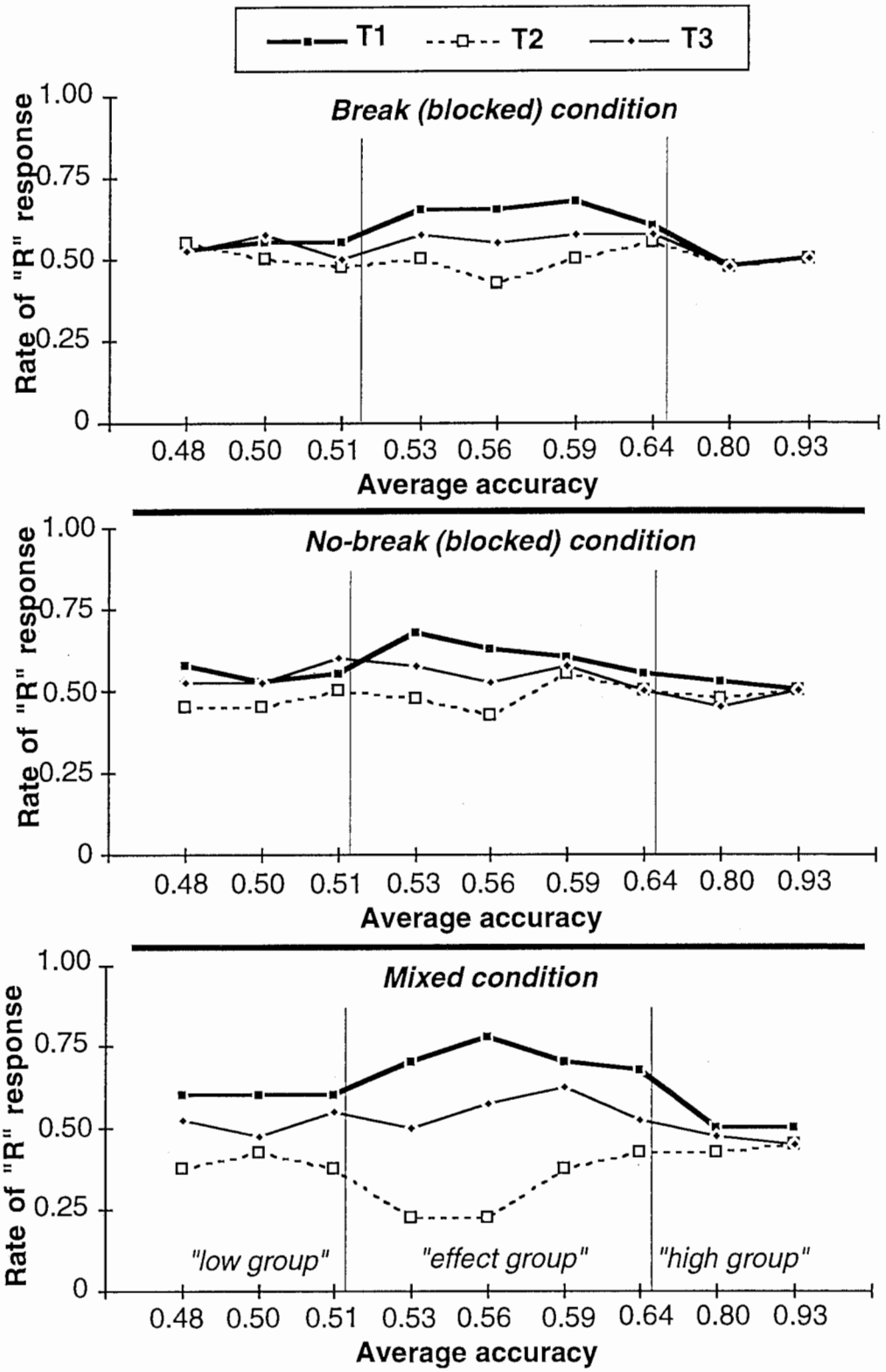


Figure 5: Rate of "R" response to each talker by average accuracy in blocked and mixed talker conditions in Experiment 1. Each point represents the average rate of "R" response for four subjects, with the exception of the points for subjects with average accuracy of .93, which are the averages for two subjects. Note that the numbers on the X axis are labels only.

		BLOCKED		MIXED	Change in fit:	
		Break	No-break			
GROUP	Cue	r^2	r^2	r^2	mixed-break	mixed-no-break
ALL	F1	.002	0	.018	.016	.018
	F2	.058	.077	**.113**	**.055**	.036
	F3	**.311**	**.365**	**.311**	0	**-.054**
	Duration	.036	.054	**.209**	**.173**	**.155**
	Vowel		.001	0	-.002	-.001
	Height	.003	.002	.001	-.002	-.001
EFFECT	F1	.002	.001	.032	.030	.031
	F2	.080	.095	**.122**	.042	.027
	F3	**.190**	**.218**	**.177**	-.013	-.041
	Duration	.087	.089	**.339**	**.252**	**.250**
	Vowel	.001	0	.002	.001	.002
	Height	.001	0	.001	0	.001
HIGH	F1	.001	0	0	-.001	.000
	F2	.027	.041	.036	.009	-.005
	F3	**.732**	**.753**	**.719**	-.013	-.034
	Duration	.005	.001	0	-.005	-.001
	Vowel	0	.001	0	0	-.001
	Height	0	.001	0	0	-.001
LOW	F1	0	.002	.011	.011	.009
	F2	.003	.009	.033	.030	.024
	F3	.000	.008	.003	.003	-.005
	Duration	.016	.045	**.115**	**.099**	**.070**
	Vowel	.006	.004	.006	0	.002
	Height	.007	.013	.008	.001	-.005

Table 4: Results of simple regressions of cue values to rate of "R" response for each accuracy group in Experiment 2. Values of r^2 greater than .100 are presented in boldface. Change in fit (r^2) is presented in the right-most column, and changes greater than .050 are presented in boldface.

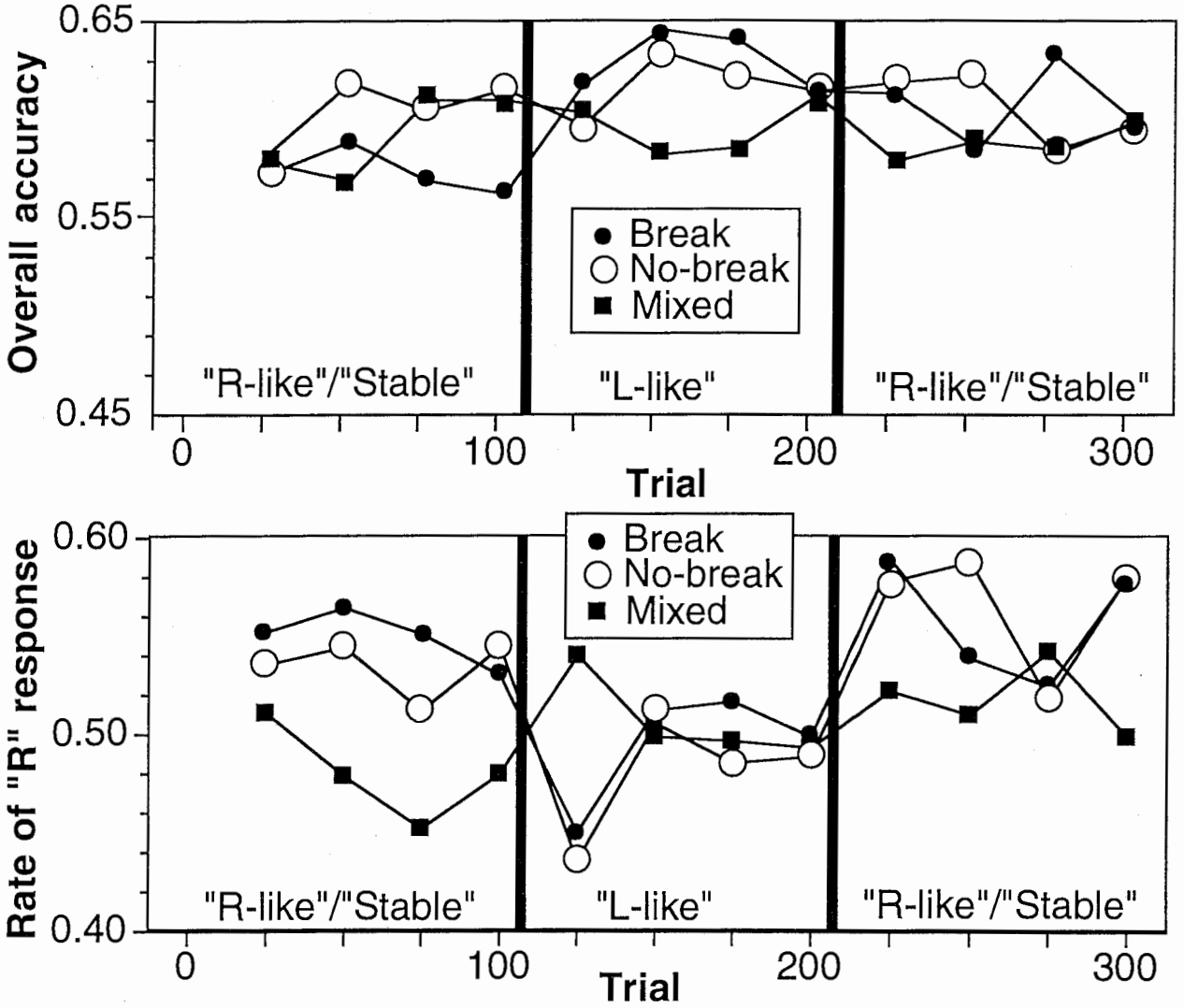
A simple regression was computed for the relation of consonant identity to F3: $\underline{y} = -.001\underline{x} + 3.152$, $\underline{r}^2 = .852$. The next highest match was with F2, although it was much lower than it was for the stimuli used in Experiment 1: $\underline{y} = -.001\underline{x} + 2.04$, $\underline{r}^2 = .024$. The mean measured values for F2 were 1090 Hz for /l/ stimuli and 1043 Hz for /r/ stimuli. The ranges were more similar than for the stimuli used in Experiment 1: 753 Hz to 1540 Hz for /l/ stimuli, and 720 Hz to 1475 Hz for /r/ stimuli. As in Experiment 1, F1 and duration were not reliable cues ($\underline{r}^2 = .0004$ and .005, respectively).

The results of simple regressions between rate of "R" response and the six cues are presented by talker condition for all subjects, and the effect, high, and low groups in Table 4. As was the case for the correlations reported for Experiment 1, these are based on correlations between R-rate and the various measures across stimuli and talkers. When correlations were computed separately for each talker, there were no substantial changes between blocked and mixed conditions for any of the talkers, except for changes in F3 which were coincident with substantial accuracy differences.

Returning to Table 4, the two cues with the highest fit to /r/-/l/ identity, F3 and F2, also had the highest fit to subjects' rate of "R" response in both talker conditions. It also appears that subjects in the effect group relied on the duration of the /r/ or /l/ portion of stimuli in the mixed-talker condition ($\underline{r}^2 = .339$) much more than in the blocked-talker conditions ($\underline{r}^2$ was approximately .088 in break and no-break conditions, thus the change in $\underline{r}^2$ was approximately .251). As in Experiment 1, the large increase in the fit to duration between the blocked and mixed conditions for the effect group suggests that subjects may have used duration as a cue to apply across talkers in the mixed-talker condition.

Tuning and design considerations. We also found the expected "tuning effect" in R-rate: when the talker changed from relatively R-like to relatively L-like, R-rate dropped. It gradually recovered, but then increased suddenly when the talker changed to a relatively R-like talker, followed by another recovery.

In Figure 6, we have plotted the 25-trial averages of overall accuracy for the three conditions in the upper panel, and rate of "R" response in the lower panel. In the 2 blocked conditions, there was a talker change every 100 trials. In the mixed condition, the talker varied



(each point represents the average of the preceding 25 trials)

Figure 6: Accuracy and rate of "R" response by trial in Experiment 2. Note that the solid vertical lines after trials 100 and 200 correspond to talker changes in the two blocked conditions (break and no-break). The data from the mixed condition are presented for comparison, but the probability of the talker changing was the same for all pairs of trials.

randomly from trial to trial. Note the decrease in "R" responses when the talker changes from "R-like/stable" to "L-like" and the increase when the talker changes from "L-like" to "R-like/stable", and the relative lack of changes in overall accuracy correlated with talker changes.

In our previous report, we discussed the following two caveats in a footnote:

> First, if we plot averages for fewer than 25 trials, the increase-decrease patterns become obscured by other increases and decreases. Second, note that there is a final increase for the two blocked conditions in the final 25 trials that we cannot explain, and which makes the previous increase-decrease patterns suspect. However, there are two points we can make in defense of Figure [6] (in addition to the comparison to the time-course of the accuracy data). First, although Kato and Kakehi (1988) and Kakehi (1992) report clear, trial-by-trial tuning, they obtained their clear curves by averaging data for 100 subjects; given our relatively small number of subjects, the trends present in Figure [6] are worthy of discussion, if not complete confidence. Second, regarding the points representing the final 25 trials in the two blocked conditions: note that the series of points for the last 100 trials is very similar to that for the first 100 trials. The final increase may reflect criterion shifts to compensate for the decrease between trials 225 and 275.

Since that time, we have realized that there was a minor flaw in the experimental design. Because stimuli were randomized in sets of 100, for some data points for some subjects, the numbers of /r/ and /l/ stimuli were unbalanced. However, even if data points based on series in which the numbers of /r/ and /l/ stimuli were unbalanced (more than 15 of one), the pattern in Figure 6 is not affected. Please note that for this sort of design, it is vital that the number of stimuli in each category be controlled within smaller series of trials. For example, we could have controlled the number of /r/s and /l/s within every four trials. This is the approach we have taken with subsequent studies.

DISCUSSION

The new analysis of subject differences allowed us to identify groups of subjects with different strategies for /r/-/l/ decisions. The knowledge that subjects with relatively low average accuracy that is somewhat above chance adopt session-specific criteria for /r/-/l/ identification may prove useful for non-native contrast training (see Magnuson and Yamada, to

appear, for some preliminary studies of the effects of talker variability on /r/-/l/ training).

The change in correlation between R-rate and duration between blocked and mixed talker conditions provides further support for our hypothesis that our subjects were attempting to find session specific criteria for /r/-/l/ decisions. In the blocked-talker condition, we were unable to find substantial correlations between R-rate and any of our physical measures (with the exception of F3, which, as the one ecologically-valid cue, is always correlated with accuracy). In the mixed-talker condition, the correlation of R-rate and duration of the initial /r/ or /l/ portion of our stimuli increased substantially. This suggests that subjects may have chosen duration as a cue that could be applied across talkers. In addition, we were not able to find such a change in correlation when the talkers were considered separately, which confirms further that subjects were using cues across talkers -- that is, they were setting session-specific criteria with possibly little regard for talker differences.

Finally, we urge other experimenters to heed our warning regarding the need to balance numbers of stimuli in small series in multiple-category forced choice designs when analyses of change over time are desired.

REFERENCES

Joos, M. (1948). Acoustic Phonetics. Supplement to Language, 24. Baltimore: Waverly Press.

Kakehi, K. (1992). Adaptability to differences between talkers in Japanese monosyllabic perception. In Y. Tohkura, Y. Sagisaka, and E. Vatikiotis-Bateson (Eds.), Speech Perception, Speech Production, and Linguistic Structure, pp. 135 142. Tokyo: OHM.

Kato and Kakehi, K. (1988). Listener adaptability to individual speaker differences in monosyllabic speech perception. Journal of the Acoustical Society of Japan, 44, 180-186.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. Journal of the Acoustical Society of America, 94, 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R.A., Tohkura, Y., and Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. Journal of the Acoustical Society of America, 96, 2076-2087.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. Journal of the Acoustical Society of America, 89, 874-886.

Magnuson, J. S., and Yamada, R. A. (1994.12.6). The effects of talker variability on the perception of American English /r/ and /l/ by Japanese listeners. Advanced Telecommunications Research Human Information Processing Research Laboratories Technical Report TR-H-110.

Magnuson, J. S., and Yamada, R. A. (to appear). Controlling talker variability in non-native speech contrast training. To appear in the Proceedings of the 1995 International Congress of Phonetic Sciences.

Magnuson, J. S., Yamada, R. A., and Nusbaum, H. C. (1994). Are representations used for talker identification available for talker normalization? Proceedings of the 1994 International Conference on Spoken Language Processing, 1171-1174.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. Journal of the Acoustical Society of America, 85, 2088-2113.

Nusbaum, H. C., and Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka, and E. Vatikiotis-Bateson (Eds.), Speech Perception, Speech Production, and Linguistic Structure, pp. 113-134. Tokyo: OHM.

O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., and Cooper, F. S. (1957). Acoustic cues for the perception of initial /w, r, l/ in English. Word, 13, 25-43.

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 309-328.

Peterson, G. E. and Barney, H. L. (1952). Control methods used in a study of vowels. Journal of the Acoustical Society of America, 24, 175-184.

Underbakke, M., Polka, L., Gottfried, T. L., and Strange, W. (1988). Trading relations in the perception of /r/-/l/ by Japanese learners of English. Journal of the Acoustical Society of America, 84, 90-100.

Yamada, R. A. (1993). Effect of extended training on /r/ and /l/ identification by native speakers of Japanese. Journal of the Acoustical Society of America, 93, 2391 (abstract).

Yamada, R. A., and Tohkura, Y. (1990). Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese. Proceedings of the 1990 International Conference on Spoken Language Processing, 757-760.

Yamada, R. A., and Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. Perception & Psychophysics, 52, 376-392.

AUTHOR NOTES

Preliminary results from Experiments 1 and 2 were reported at the 127th meeting of the Acoustical Society of America (Magnuson, J. S. and Yamada, R. A. [1994a]. Talker variability and the identification of American English /r/ and /l/ by Japanese subjects. Poster presented at the 127th meeting of the Acoustical Society of America, Cambridge, MA, USA, 7 June 1994. Journal of the Acoustical Society of America, 95, 2872). Preliminary results from Experiment 3 were presented at the 1994 Fall meeting of the Acoustical Society of Japan (Magnuson, J. S., and Yamada, R. A. [1994b]. The effects of talker variability on the perception of American English /r/ and /l/ by Japanese subjects: Normalization or criteria-setting? Proceedings of the Fall 1994 Meeting of the Acoustical Society of Japan, 357-358).

We owe a great debt to Dr. Yoh'ichi Tohkura, whose suggestions and insight inspired the analyses presented in this report. We also thank Prof. David Pisoni for providing us with the recordings of the talkers; Prof. Winifred Strange, Prof. Kevin Munhall, Prof. Howard Nusbaum, Dr. Hideki Kawahara, Inge-Marie Eigsti and John S. Pruitt for comments which substantially improved this paper; Takahiro Adachi for programming; and Rie Kawakami and Chiemi Inoue for testing subjects.