Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*)

Andrew J. Lotto,[a] Keith R. Kluender, and Lori L. Holt
University of Wisconsin, Madison, Wisconsin 53706

(Received 18 April 1996; revised 19 March 1997; accepted 16 April 1997)

When members of a series of synthesized stop consonants varying in third-formant ($F3$) characteristics and varying perceptually from /da/ to /ga/ are preceded by /al/, human listeners report hearing more /ga/ syllables than when the members of the series are preceded by /ar/. It has been suggested that this shift in identification is the result of specialized processes that compensate for acoustic consequences of coarticulation. To test the species-specificity of this perceptual phenomenon, data were collected from nonhuman animals in a syllable ''labeling'' task. Four Japanese quail (*Coturnix coturnix japonica*) were trained to peck a key differentially to identify clear /da/ and /ga/ exemplars. After training, ambiguous members of a /da/–/ga/ series were presented in the context of /al/ and /ar/ syllables. Pecking performance demonstrated a shift which coincided with data from humans. These results suggest that processes underlying ''perceptual compensation for coarticulation'' are species-general. In addition, the pattern of response behavior expressed is rather common across perceptual systems. © *1997 Acoustical Society of America.* [S0001-4966(97)01608-1]

PACS numbers: 43.71.An, 43.71.Es, 43.80.Lb [WS]

INTRODUCTION

One of the wonders of speech communication is the remarkable symmetry between perception and production. Articulatory dynamics and constraints shape the resultant waveform, and in many cases, perceptual processes of the listener act in ways that appear to respect these constraints and dynamic properties. The beauty of this symmetry (or synergy) has motivated several comprehensive theories of speech communication. For example, in the revised Motor Theory of Liberman and Mattingly (1985), symmetry is said to arise because of the shared currency of speech perception and production: gestural representations. Auditorist theories, such as espoused by Diehl and his colleagues (e.g., Diehl and Kluender, 1989; Diehl *et al.*, 1991; Kingston and Diehl, 1995), suggest that much of the symmetry is a result of talkers producing speech sounds in a manner that exploits general operating characteristics of the auditory system. Direct Realist approaches (e.g., Fowler, 1986, 1996) account for such symmetry in terms of perceptual ''recovery'' of vocal-tract dynamics from the acoustic waveform. Despite this diversity of theoretical accounts, it is clear that effective communication requires perception and production to be in relatively close accord.

One case of symmetry between speech perception and production that has received a considerable amount of empirical attention is the effect of a preceding liquid on stop-consonant perception. Mann (1980) presented listeners with members of a series of synthesized consonant–vowel (CV) syllables varying perceptually from /da/ to /ga/ preceded by natural utterances of either /al/ or /ar/. Subjects identified the CVs as /ga/ more often following /al/ than following /ar/. This shift in responses is complementary to the coarticulatory influences on CV production following /al/ and /ar/. Due to the assimilative nature of coarticulation, the place of vocal-tract occlusion of a stop consonant is more anterior following /al/ than following /ar/. Because the alveolar stop /d/ is produced at an anterior place in the oral cavity, and the velar stop /g/ is produced with a posterior occlusion, /al/ productions result in subsequent CVs being more /da/-like articulatorally and acoustically, while CVs following /ar/ are more /ga/-like. Thus subjects' perceptual responses seem to compensate for acoustic effects of coarticulation; more /ga/ identifications result for CVs following /al/.

Subsequent research has demonstrated this context effect on CV identification for Japanese listeners who are ineffective in distinguishing /al/ and /ar/ (Mann, 1986) and for four-month-old infants (Fowler *et al.*, 1990). These results have been attributed to perceptual representations of, or recovery of, specific vocal-tract dynamics and constraints. According to Mann (1980), ''...speech perception must somehow operate with tacit reference to the dynamics of speech production and its acoustic consequences.''

However, this theoretical view was strongly challenged by results reported by Lotto and Kluender (in press). Participating in a forced-choice identification task like that employed by Mann (1980), subjects identified members of a CV series as /ga/ more often following /al/ than when following /ar/, even when the two syllables were produced by very physically different talkers of opposite gender. Furthermore, the effect remained when the preceding context was a sine-wave caricature modeling a very limited aspect ($F3$ transition) of the /al/ and /ar/ syllables. Lotto and Kluender suggest that ''perceptual compensation for coarticulation'' is probably not due to knowledge of or recovery of specific vocal-tract dynamics because the context effect remains robust for contextual sounds which were clearly not produced by a similar

[a]Correspondence to: Andrew J. Lotto, Department of Psychology, University of Wisconsin, 1202 West Johnson Street, Madison, WI 53706, (608) 262-6110, ajlotto@facstaff.wisc.edu

vocal tract or, indeed, by any vocal tract. They propose that the effect is of general auditory origin and describe the pattern of results as frequency contrast. Redescribed in this light, the results are: following a syllable with a high $F3$-offset frequency (/al/), more low-frequency-$F3$-onset identifications (/ga/) result, and following a syllable with a low $F3$-offset frequency (/ar/), more high-frequency-$F3$-onset identifications (/da/) are obtained. In light of this general framework, Lotto and Kluender (in press) presented subjects with CV syllables similar to those used by Mann (1980) preceded by constant-frequency sine-wave tones with frequencies equivalent to $F3$-offset frequencies of natural /al/ and /ar/ syllables which served as context in the Mann study. The resulting shift in identification functions was actually slightly larger than that obtained by Mann (1980) for natural-speech contexts.

One may assume that accounts which rely on speech-specific mechanisms to accommodate coarticulatory influences would have difficulty explaining these nonspeech results. However, there remain concerns about the validity of research with nonspeech analogs as critical experiments for deciding issues of speech-specificity. Kuhl (1978, 1986a, b) has discussed the possibility that nonspeech stimuli may be accommodated by speech-specific mechanisms with rather broad application. Processes which rely upon abstract kinematic consequences of articulatory dynamics have been proffered, for example, to explain the finding that some subjects are able to hear complexes of sine waves as speech (Remez *et al.*, 1981; Remez *et al.*, 1994).

Even if one does not accept that mechanisms have evolved exclusively for perceiving speech, it is possible that the overlearned nature of speech can affect the perception of quasi-periodic nonspeech sounds. Perceptual heuristics[1] for managing the kinematic characteristics of speech may be developed through the near-constant exposure to these sounds. As a result, contextual sine waves may be processed in a speechlike manner.

In order to test the generality of this context effect and to determine if an account based on general ''frequency contrast'' is viable, an experiment was designed exploiting a nonhuman animal model of speech perception. Previous animal studies of speech perception have been used to assess general auditory processes without confounds of effects of experience and unencumbered by purported speech-specific processes (e.g., Kuhl and Miller, 1975, 1978; Kluender, 1991; Kluender and Lotto, 1994; Dooling *et al.*, 1995). Animals are unlikely recipients of innate speech-specific mechanisms. Consequently, analogous performance on speech tasks for animals and humans, together with the virtue of parsimony, discourages accounts of speech perception which rely on innate representations of gestural dynamics.[2]

I. EXPERIMENT

Four Japanese quail (*Coturnix coturnix japonica*) served as subjects in an experiment designed to test the species-specificity of the contextual effects reported in Mann (1980). Japanese quail have been used successfully in previous experiments concerning the complementary nature of speech perception and production (e.g., Kluender *et al.*, 1987; Kluender, 1991; Kluender and Lotto, 1994). These birds have shown the ability to ''identify'' CV syllables varying in information specifying place of articulation through differential pecking (Kluender *et al.*, 1987). The present experiment tested whether their CV ''labeling'' would be affected by intersyllabic context.

A. Methods

1. Subjects

Four adult Japanese quail served as subjects in the labeling experiment. Free-feed weights ranged from 123 to 154 g.

2. Stimuli

Stimuli were identical to those from the synthesized-speech condition of experiment 2 from Lotto and Kluender (in press). A ten-step series of CV syllables (/da/-/ga/) varying in $F3$-onset frequency was synthesized using the cascade synthesizer described in Klatt (1980). End-point stimuli were based on natural productions of a male talker speaking the syllables in isolation. For these CVs, onset frequency of $F3$ varied from 1800 to 2700 Hz in 100 Hz steps. Then, from onset, $F3$ frequency changed linearly to a steady-state value of 2450 Hz over an 80-ms transition.[3] Amplitude of $F3$ was approximately 3 dB less intense than $F1$ and $F2$ at onset. All other synthesizer parameters remained constant across members of the series. Frequency of the first formant ($F1$) increased from 300 to 750 Hz and second-formant ($F2$) frequency decreased from 1650 to 1200 Hz over 80 ms. Fundamental frequency ($f0$) was 110 Hz from onset until decreasing linearly to 95 Hz over the last 50 ms. Total stimulus duration of synthesized CVs was 250 ms.

Three syllables serving as preceding context were also synthesized with values based on utterances of the same talker upon whose productions the CV series was modeled. All three syllables were 250 ms in duration and had a constant $f0$ of 110 Hz. Two of the preceding-context syllables were synthesized versions of /al/ and /ar/. For each syllable, the first 100 ms consisted of the same steady-state vowel. The frequencies of the first four formants during this steady state were 750, 1200, 2450, and 2850 Hz, respectively. Following this 100-ms vowel were 150-ms linear formant transitions. Offset frequencies for the first four formants in the /al/ syllable were 564, 956, 2700, and 2850 Hz, respectively. For the /ar/ syllable these values were 549, 1517, 1600, and 2850 Hz. The third preceding context was a 250-ms steady-state vowel /a/ synthesized with the same parameters as for the vowel in each VC.

Stimuli were synthesized with 12-bit resolution at a 10-kHz sampling rate, matched in rms level and stored on computer disk. Stimulus presentation was under control of an 80386 computer. Following D/A conversion (Ariel DSP-16), stimuli were low-pass filtered (4.8-kHz cutoff frequency, Frequency Devices #677), amplified, and presented to subjects via a single 13-cm speaker (Peerless 1592) in a tuned enclosure providing flat frequency response from 40 to 5000 Hz. Sound level was calibrated by placing a small sound-level meter (Bruel & Kjaer 2232) in the chamber with the microphone positioned at approximately the same height and distance from the speaker as the performing bird's head.

3. Procedure

Quail first were trained by means of operant procedures to discriminate stimuli from each end of the CV series. For two birds (quail 1 and quail 2), CVs with low $F3$-onset frequencies (1800 and 1900 Hz) signaled positive reinforcement (/ga/-positive), while for the other two birds (quail 3 and quail 4), CVs with high $F3$-onset frequencies (2600 and 2700 Hz) were reinforced (/da/-positive). Following 18 to 22 h of food deprivation (adjusted to each bird individually for optimal performance[4]), birds were placed in a soundproof operant chamber (Industrial Acoustics Corp., model AC1) inside a larger single-wall soundproof booth (Suttle Acoustics Corp.). In a go/no-go identification task, birds pecked a single lighted 1.2-cm-square key located 15 cm above the floor and centered below the speaker. Stimuli were presented, responses were recorded, and reinforcement was controlled by an 80386 microcomputer.

For three quail, the training sequence was identical. CVs were presented in the three contexts (/a/, /al/, and /ar/) from the beginning of training. For the other quail (quail 4), CVs were presented in isolation until the bird's peck ratios reached a performance criterion (10:1 peck ratio for positive versus negative stimuli), then training was continued with inclusion of contextual sounds. This alternative procedure for quail 4 resulted in no discernible difference in the final data.

During training with contextual sounds, stimuli consisted of a disyllable including one of the three contexts (/a/, /al/, or /ar/) followed by a 50-ms silent interval (typical of natural productions) and then one of the four training CV syllables (with $F3$ frequencies of 1800, 1900, 2600, and 2700 Hz). Appending of syllables was accomplished digitally online during the experiment. On each trial, a disyllable was repeatedly presented once per 1550 ms at an average peak level of 70 dB SPL. (During the initial training of quail 4, single CVs were repeated once per 1550 ms.) On a trial-by-trial basis, the intensity of the disyllable (or single CV) was varied randomly from 70 dB by $\pm 0-5$ dB [mean=70 dB SPL] through a computer-controlled digital attenuator (Analog Devices 7111). Average duration of each trial was 30 s, varying geometrically from 10 to 65 s. Intertrial interval was 15 s. Responses to positive stimuli were reinforced on a variable-interval schedule by 1.5–2.5 s access to food from a hopper beneath the peck key. Duration of reinforcement was also adjusted for each bird for consistent performance. Average interval to reinforcement was 30 s (10–65 s), so that positive stimuli were reinforced on an average of once per trial. Note that when a trial was long (e.g., 57 or 65 s) and times to reinforcement were short (e.g., 10 or 12 s), reinforcement was available more than once. Likewise, on shorter positive trials reinforcement did not become available if time to reinforcement was longer than the trial. Any reinforcement interval that did not expire during one positive trial carried over to the next positive trial. Such intermittent reinforcement encouraged consistent peck rates during later nonreinforced testing trials. During negative trials, birds were required to refrain from pecking for 5 s for presentation of the stimulus, and the trial to be terminated. This procedure has been used successfully to train Japanese quail in similar speech perception tasks (Kluender, 1991; Kluender and Lotto, 1994).

Following magazine training and autoshaping procedures, reinforcement contingencies were gradually introduced over a 1-week period in sessions of 60 to 72 trials each. During that first week the average amplitude of the stimuli was increased from 50 to 70 dB SPL in order to introduce sound without startling the birds. (Following the attainment of a performance criterion, contextual sounds were gradually presented to quail 4. The average amplitude of these context sounds was increased from about 40 to 70 dB over a 2-week period.) Also during this first week: average trial duration increased from 5 to 30 s; intertrial interval decreased from 40 to 15 s; average time to reinforcement was increased from 5 to 30 s; access to the food hopper was decreased from 4.0 to 2.0 s; and ratio of positive to negative trials decreased from 4:1 to 1:1.

All birds learned quickly to respond differentially to high $F3$-onset frequency versus low $F3$-onset frequency CVs, pecking at least twice as often to positive stimuli versus negative stimuli by the end of 50 days of training (3600 trials). One bird (quail 2) was pecking at a 2:1 ratio after only 23 days of training (1656 trials). Whereas Kluender and Lotto (1994) found 2:1 performance after only 20 days of training for a voiced/voiceless distinction, the task for the current birds may be considered more difficult in that reinforced discrimination must be made solely on a small change in relative amplitudes of harmonics as opposed to the multifarious cues which result from changes in voice-onset time. Birds continued to train with the four extreme CVs in three contexts until they achieved 10:1 performance for positive versus negative stimuli. Attaining this level of performance required between 140 and 163 days of training (10 080 to 11 160 trials). (In addition, quail 4 needed 88 days (6336 trials) to reach 10:1 performance after contextual sounds were added.)

Following training, the four birds were tested on novel CV syllables with intermediate $F3$-onset frequencies ranging from 2000 to 2500 Hz (100-Hz steps) in all three contexts (/a/, /al/, and /ar/). Due to uncontrollable scheduling conflicts, the number of testing trials with each novel disyllable varied between birds. Quail 1 was presented all six intermediate CVs with each context eight times across 16 days of testing. Quail 2 responded to each disyllable 10 times over 20 days; nine times over 18 days for quail 3; 20 times over 40 days for quail 4. During a single test session, nine of the possible 18 novel disyllables (six CVs×three contexts) were presented, each on one 30-s trial. During the presentation of novel disyllables, no contingencies were in effect. Birds neither received food reinforcement nor needed to refrain from pecking for presentation to terminate after 30 s. Each testing session of 69 total trials began with 15 reinforced trials with training disyllables followed by nine nonreinforced trials with novel disyllables interspersed amongst 45 reinforced

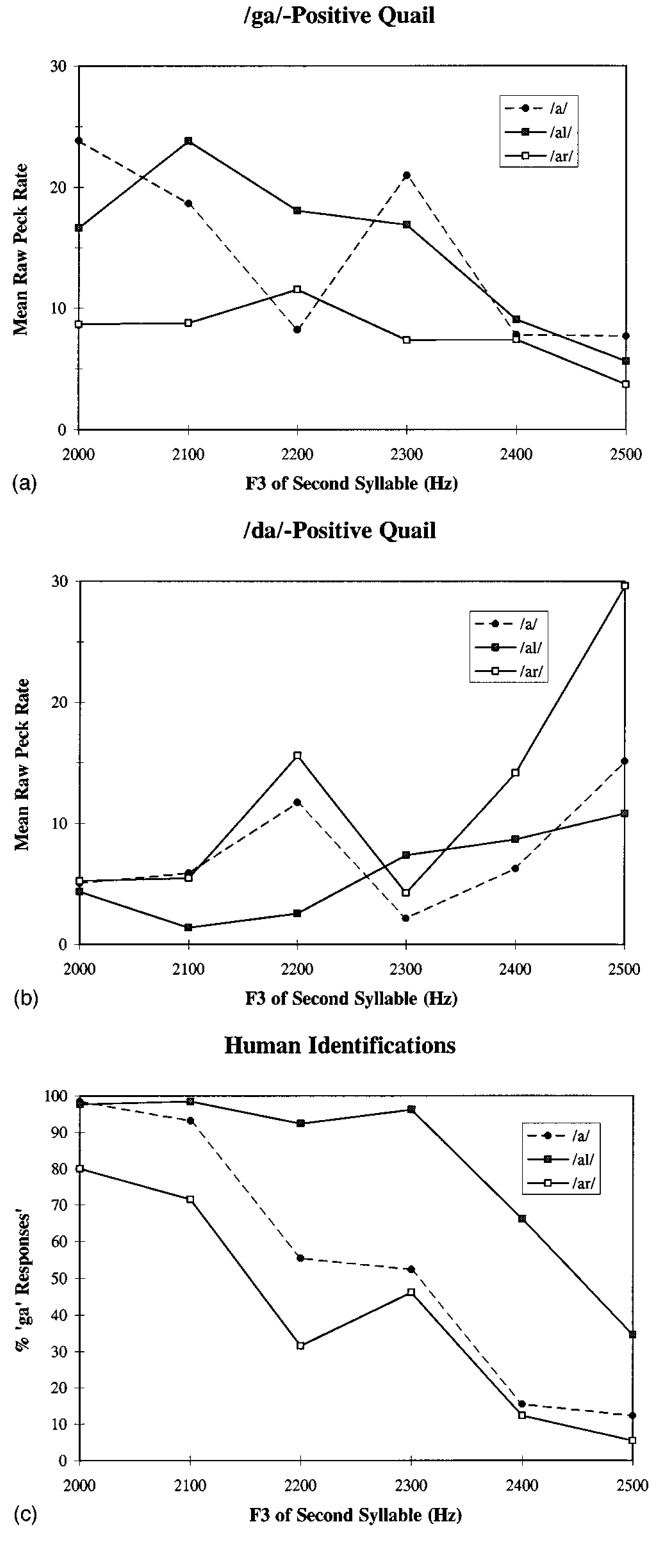

FIG. 1. Mean raw peck rates for each of the six novel /Ca/ stimuli in each of the three context conditions. (a) Average peck rates for birds trained to peck to /ga/ stimuli. (b) Average peck rates for birds trained to peck to /da/ stimuli. (c) Mean percentage of /ga/ identifications for human listeners from Lotto and Kluender (in press).

trials with training stimuli. Novel trials could not occur until 15 training stimuli trials had been presented to assure that each bird ''settled into'' the task before responding to test stimuli. For all training and testing sessions, stimuli were randomly ordered for each bird.

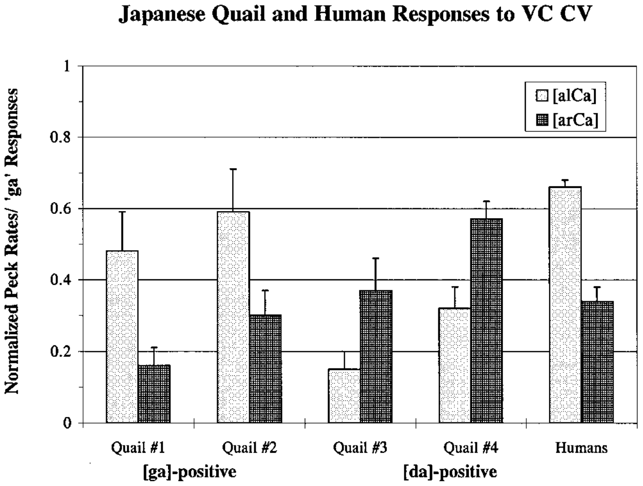

FIG. 2. Histogram contains normalized peck rates (percentages of range of mean peck rates) along with attendant standard errors for intermediate CV syllables in the context of /al/ and /ar/. Human data is from Lotto and Kluender (in press) synthesized-speech condition of experiment 2. Represented here are the percent /ga/ responses to those CVs which served as test stimuli in the current quail experiment. Mean percentages were normalized so that they summed to 1.00 to put them on a scale similar to that used for peck rates.

B. Results and discussion

For each bird, raw pecks were collected for each test trial. These test trials were all fixed at a duration of 30 ms. Figure 1 displays the raw peck data for each of the six novel stimuli in each of the three context conditions. Figure 1(a) is a plot of the data averaged across the responses of the two birds trained to peck to /ga/ (low-frequency $F3$) stimuli and Fig. 1(b) is a plot of the data collected from the birds trained to peck to /da/ (high-frequency $F3$) stimuli. Figure 1(c) is a plot of human identification functions from Lotto and Kluender (in press) for the novel stimuli. Note that there is a nonlinearity for the birds and humans for the CV with an $F3$ onset of 2200 Hz. This stimulus differs only in $F3$-onset frequency from the other stimuli and experiments are being conducted to discover the reason for this nonlinearity. It is interesting that there is such qualitative agreement in the functions for two different species, presumably performing different tasks.

An alternative data representation was also calculated to deal with the inherent variance arising from the different peck rates of individual birds. Total pecks to each disyllable were summated for each ''run'' through all of the novel stimuli, i.e., the data for all 18 novel disyllables collected across two days. Mean peck rates (pecks per 30-s trial) were calculated for each of the three contexts (/a/, /al/, /ar/). These means were then transformed into percentages of the mean range. The lowest mean for each bird was subtracted from each mean and the result was divided by the range of mean scores (maximum mean−minimum mean). For example, for quail 3, the highest mean peck rate for any context was 39.33 pecks/30 s. This was the mean peck rate across all novel stimuli with the /ar/ context on its first testing run. The minimum mean peck rate for this bird was 0.0 pecks/30 s for all novel stimuli in the /al/ context on the sixth run. The nor-

TABLE I. Results of t tests and Wilcoxon signed-rank tests on normalized peck rates for each subject.

Bird	/al/ mean	/ar/ mean	t value	p value	Wilcoxon	p value
Quail 1	0.48	0.16	3.57	0.009	0.0	0.008
Quail 2	0.59	0.30	2.50	0.034	8.0	0.047
Quail 3	0.15	0.37	2.74	0.025	4.0	0.028
Quail 4	0.32	0.57	6.41	0.000	6.0	0.000

malized mean peck rate for the /ar/ context of the third run would be

$$\frac{(\text{mean for /ar/ of third run})-(\text{minimum mean})}{(\text{maximum mean})-(\text{minimum mean})}$$

$$=\frac{(20.83)-(0.00)}{(39.33)-(0.00)}=0.53.$$

This transform normalizes the mean peck rates in terms of the range of peck rates for an individual bird, thus minimizing variance that arises from the fact that some birds are ''heavier'' peckers than others. It has been determined through Monte Carlo simulations that such a range transform increases power without increasing the likelihood of type I error (Bush *et al.*, 1993). Normalized mean peck rates with attendant standard errors for the /al/ and /ar/ contexts are presented in Fig. 2. Quails 1 and 2 were trained to peck to /ga/ stimuli and show increased pecking in the /al/ condition. Conversely, quails 3 and 4, which were trained to peck to the /da/ end points, pecked more to novel stimuli following /ar/. For comparison, data from adult humans hearing the same stimuli (Lotto and Kluender, in press) are presented. These data were normalized by dividing the number of /ga/ responses in each condition by the total number of /ga/ responses across conditions.

Matched-pairs t tests were computed for the difference between normalized peck rates in /al/ and /ar/ contexts for each bird. Runs through the stimuli (i.e., 2-day sessions in which each stimulus is heard once) were treated as the random variable for the tests. For each bird, normalized peck rates were significantly different in /al/ than in /ar/ contexts.[5] Table I presents the outcomes of each t test along with the outcomes of independently conducted nonparametric Wilcoxon signed-rank tests. For birds trained to peck in response to CVs with high $F3$-onset frequencies (/da/-positive: quail 3 and quail 4), peck rates increased for CV syllables following /ar/. For birds trained to peck to CVs with low $F3$-onset frequencies (/ga/-positive: quail 1 and quail 2), peck rates increased for CV syllables following /al/. This pattern of responses is analogous to that of human listeners in identification tasks, who respond with /ga/ identifications more often following /al/ than following /ar/ (Mann, 1980, 1986; Fowler *et al.*, 1990; Lotto and Kluender, in press).

II. GENERAL DISCUSSION

The experiment in this report was designed to test the generality of the effect of preceding liquid identity on stop-consonant identification first reported in Mann (1980). The fact that nonhuman animals demonstrated effects of context on labeling of CV syllables similar to that found for human listeners suggests that this effect is of general auditory origin. Japanese quail are unlikely to possess processes designed specifically for the domain of human speech. These birds also had no chance to learn the covariation of formant frequencies caused by assimilative processes of coarticulation. As a matter of parsimony, one is thrust into the theoretical position to recommend that ''perceptual compensation for coarticulation'' is the outcome of a rather general auditory process.

Consequences of this general process can be described as contrastive. Quail trained to peck to CVs with *low* $F3$-onset frequencies (/ga/-positive) pecked more to intermediate values of $F3$-onset frequency when CVs were preceded by a syllable with *a high*-frequency $F3$ offset (/al/). Quails trained to peck to CVs with *high* $F3$-onset frequencies (/da/-positive) pecked more to intermediate values of $F3$-onset frequency when CVs were preceded by a syllable with a low-frequency $F3$ offset (/ar/). This contrastive pattern has now been shown for adult human English speakers (Mann, 1980; Lotto and Kluender, in press), adult human Japanese speakers (Mann, 1986), 4-month-old infants (Fowler *et al.*, 1990), and Japanese quail.

In fact, contrastive perceptual effects, as those noted here for human and bird listeners, are epidemic. In the visual system, perceptual contrast has been described for lightness perception (Koffka, 1935; Wallach, 1948), line orientation (Gibson, 1933, 1937; Gibson and Radner, 1937), size, position, and curvature (Kohler and Wallach, 1944), spatial frequency (Blakemore and Sutton, 1969), depth (Ames, 1935; Kohler and Emery, 1947; Bergman and Gibson, 1959), and color (Cathcart and Dawson, 1928–1929). Contrastive effects have been witnessed in tempo of behavior (Cathcart and Dawson, 1927–1928) and lifting of weights (Guilford and Park, 1931; Sherif *et al.*, 1958). In audition, frequency contrast has been demonstrated (Cathcart and Dawson, 1928–1929; Christman, 1954) as has contrast in lateralization of a sound (Flügel, 1920–1921).

In speech perception literature, subjects' responses often can be described in terms of contrast of some type. For example, identification boundaries for members of a stop/glide series varying in transition duration shift toward longer transitions (i.e., more *short*-transition responses) when syllable duration is *increased* (Miller and Liberman, 1979). Similarly, syllable-final consonants are judged more often to be voiced (*shorter* silent interval) when the duration of the preceding vowel is *increased* (Denes, 1955; Raphael, 1972; Port and Dalby, 1982). In each of these cases, including VC CV context effects described in this report, contrastive perception appears to compensate for articulatory regularities.

Given the ubiquity of contrastive perception, one may conjecture that it serves an important adaptive purpose. But what purpose is served by a general perceptual characteristic which causes, at first blush, seemingly nonveridical percepts? Why would a Japanese quail benefit from a process which alters perceived frequency of spectral components depending on context?[6] Part of the answer may lie in the remarkable symmetry between speech perception and production that was noted in the Introduction.

Due to the variables of inertia and mass, physical systems tend to be assimilative. The configuration of a system at time t is significantly constrained by its configuration at time $t-1$. The set of possible transformations from time t to $t-1$ is also limited (e.g., by a constraint of rigidity; Ullman, 1984). Perceptual systems have developed in an environment governed by particular physical laws and it is probable that perceptual processes respect these laws. This is a form of an argument advanced by Shepard (1984) under the name ''Psychophysical Complementarity.''

Because very rapid change is the exception for physical systems with mass and inertia, signs of change are emphasized through the processes of perceptual contrast. This is due, perhaps, to the ecological importance of rapid change, especially as it relates to the default of continuance in physical systems. For example, lightness contrast emphasizes boundaries at which there is a rapid change in luminance. This may help establish the borders of separate objects. This is similar to accounts of perceptual contrast which assume that perception is referenced to or ''anchored'' upon some previously presented standard (e.g., Helson, 1964; Warren, 1985).

As for speech communication, coarticulation is due, at least in part, to physical constraints on articulators and to dynamic variables of mass and inertia (Ostry *et al.*, 1996). Vocal-tract shape changes relatively smoothly over time. At conversational speaking rates, articulators often undershoot target articulations which are produced in clear speech (Lindblom, 1963). Because it is a physical system, signals generated by a vocal tract are perceived in a contrastive manner by both humans and Japanese quail. The resulting symmetry of production and perception is not serendipitous, but is a consequence of organisms having evolved to interact with physical systems which are constrained across time. In this light, the results with Japanese quail are not too surprising. The physical environment of the antecedents of the quail likely resemble that of early hominids and analogous perceptual solutions probably have developed.

This account is, for now, too superficial to qualify as a full theoretical account. However, it does have some concepts in common with major theories in speech communication. Along with Motor Theory, it acknowledges that articulators are highly constrained across time and that the resultant signal is largely a product of articulatory constraints and dynamics. It mirrors Direct Realism in its appreciation for the ecology of sound and the sources which produce it. It shares with Auditorist theories a tenet that general auditory processes are culpable for many of the phenomena of speech perception.

The disjunctions between this account and those which precede it must also be given consideration. It denies the necessity of tacit gestural representations and speech-specific processes as proposed by Motor Theory. It cannot be characterized as Direct Realism in as much as the proposed percepts are not veridical and need not correspond to real ''objects.'' And finally, its emphasis is shifted from traditional Auditorist accounts. In the present account, the symmetry between speech perception and production arises from the perceptual system accommodating the constraints on physical systems such as articulators, and not from the dictatorial demands of the operating characteristics of the auditory system.

As facile as this account may be, it addresses the mounting evidence of similar perceptual behavior for nonhuman animals and humans with speech stimuli (e.g., Kluender *et al.*, 1987; Kluender, 1991; Kluender and Lotto, 1994; Dooling *et al.*, 1995). In the present case, Japanese quail have shown context effects for speech sounds; an effect previously described as ''perceptual compensation for coarticulation'' (e.g., Mann 1980). As impressive as the symmetry between speech perception and production is, it is not an exclusively human achievement.

ACKNOWLEDGMENTS

This work was supported by NIDCD Grant No. DC-00719 and NSF Young Investigator Award DBS-9258482 to the second author and by Sigma Xi Dissertation Research Award to the first author. Some of the data were presented at the International Congress of Phonetic Sciences, August 1995, in Stockholm.

[1] ''Heuristic'' is intended only to refer to a class of algorithms which offer potentially fallible solutions to problems, but are nevertheless useful in most situations. The term is often expanded with theoretical content in cognitive science.

[2] Comparable perceptual behavior of humans and animals should not bother purveyors of Direct Realist theories. To the contrary, since, by this view, the information specifying articulatory dynamics is inherent in acoustic signals produced by vocal tracts, nonhuman animals should be able to recover this information (see, e.g., Fowler, 1996). However, the results of Lotto and Kluender (in press) should be particularly troubling as they demonstrate contextual effects for sounds which clearly originate from different sources.

[3] Formant transitions of 80-ms duration may seem to be rather long, but these were measured from natural productions and are comparable to the 100-ms transitions used by Mann (1980).

[4] Optimal performance was defined as the highest ratio of pecks to positive versus negative stimuli. Birds were idiosyncratic with regard to the amount of deprivation that resulted in the most stable performance, and weights ranged from 80% to 95% of free-feed weight at the time of training/testing.

[5] For all birds, mean peck rates to CVs in the /a/ context fell between those for CVs in the /al/ and /ar/ contexts.

[6] It should be noted that changes in formant frequencies are not changes in frequency, *per se*. Harmonics remain at the same frequency, but there is a change in the relative energy across the spectrum. It may be more correct to state that the perception of relative energy is context-dependent.

Ames, A. (**1935**). ''Aneiseikonia—a factor in the functioning of vision,'' Am. J. Opthamology **28**, 248–262.

Bergman, R., and Gibson, J. J. (**1959**). ''The negative aftereffect of a surface slanted in the third dimension,'' Am. J. Psychol. **72,** 364–374.

Blakemore, C., and Sutton, P. (**1969**). ''Size adaptation: a new aftereffect,'' Science **166,** 245–247.

Bush, L. K., Hess, U., and Wolford, G. (**1993**). ''Transformations for within-subject designs: A Monte Carlo investigation,'' Psychol. Bull. **113,** 566–579.

Cathcart, E. P., and Dawson, S. (**1927–1928**). ''Persistence: A characteristic of remembering,'' Brit. J. Psychol. **18**, 262–275.

Cathcart, E. P., and Dawson, S. (**1928–1929**). ''Persistence (2),'' Brit. J. Psychol. **19**, 343–356.

Christman, R. J. (**1954**). ''Shifts in pitch as a function of prolonged stimulation with pure tones,'' Am. J. Psychol. **67,** 484–491.

Diehl, R. L., and Kluender, K. R. (**1989**). ''On the objects of speech perception,'' Ecol. Psychol. **1**, 121–144.

Diehl, R. L., Kluender, K. R., Walsh, M. A., and Parker, E. M. (**1991**). ''Auditory enhancement in speech perception and phonology,'' in *Cogni-*

tion and the Symbolic Processes: Applied and Ecological Perspectives, edited by R. Hoffman and D. Palermo (Erlbaum, Hillsdale, NJ), pp. 59–75.

Denes, P. (**1955**). ''Effect of duration on the perception of voicing,'' J. Acoust. Soc. Am. **27,** 761–764.

Dooling, R. J., Best, C. T., and Brown, S. D. (**1995**). ''Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*),'' J. Acoust. Soc. Am. **97,** 1839–1846.

Flügel, J. C. (**1920–1921**). ''On local fatigue in the auditory system,'' Brit. J. Psychol. **11**, 105–134.

Fowler, C. A. (**1986**). ''An event approach to the study of speech perception from a direct-realist perspective,'' J. Phon. **14**, 3–28.

Fowler, C. A. (**1996**). ''Listeners do hear sounds, not tongues,'' J. Acoust. Soc. Am. **99,** 1730–1741.

Fowler, C. A., Best, C. T., and McRoberts, G. W. (**1990**). ''Young infants' perception of liquid coarticulatory influences on following stop consonants,'' Percept. Psychophys. **48,** 559–570.

Gibson, J. J. (**1933**). ''Adaptation, after-effect and contrast in the perception of curved lines,'' J. Exp. Psychol. **16,** 1–31.

Gibson, J. J. (**1937**). ''Adaptation with negative after-effect,'' Psychol. Rev. **44,** 222–244.

Gibson, J. J., and Radner, M. (**1937**). ''Adaptation, after-effect, and contrast in the perception of tilted lines. I. Quantitative Studies,'' J. Exp. Psychol. **20,** 453–467.

Guilford, J. P., and Park, D. G. (**1931**). ''The effect of interpolated weights upon comparative judgments,'' Am. J. Psychol. **43,** 589–599.

Helson, H. (**1964**). *Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior* (Harper & Row, New York).

Kingston, J., and Diehl, R. L. (**1995**). ''Intermediate properties in the perception of distinctive feature values,'' in *Papers in Laboratory Phonology IV*, edited by A. Arvaniti and B. Connell (Cambridge U.P., Cambridge, England).

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67,** 971–995.

Kluender, K. R. (**1991**). ''Effects of first formant onset properties on voicing judgments result from processes not specific to humans,'' J. Acoust. Soc. Am. **90,** 83–96.

Kluender, K. R., Diehl, R. L., and Killeen, P. (**1987**). ''Japanese quail can learn phonetic categories,'' Science **237,** 1195–1197.

Kluender, K. R., and Lotto, A. J. (**1994**). ''Effects of first formant onset frequency on [-voice] judgments result from general auditory processes not specific to humans,'' J. Acoust. Soc. Am. **95,** 1044–1052.

Koffka, K. (**1935**). *Principles of Gestalt Psychology* (Harcourt, Brace & World, New York).

Kohler, W., and Emery, D. A. (**1947**). ''Figural aftereffects in the third dimension of visual space,'' Am. J. Psychol. **60**, 159–201.

Kohler, W., and Wallach, H. (**1944**). ''Figural aftereffects: An investigation of visual processes,'' Proc. Am. Philos. Soc. **88**, 269–357.

Kuhl, P. K. (**1978**). ''Predispositions for the perception of speech-sound categories: A species-specific phenomenon,'' in *Communicative and Cognitive Abilities-Early Behavioral Assessment*, edited by F. D. Minifie and L. L. Lloyd (University Park Press, Baltimore), pp. 229–255.

Kuhl, P. K. (**1986a**). ''Theoretical contributions of tests on animals to the special-mechanisms debate in speech,'' Exp. Biol. **45,** 233–265.

Kuhl, P. K. (**1986b**). ''The special-mechanisms debate in speech research: Categorization tests on animals and infants,'' in *Categorical Perception*, edited by S. Harnad (Cambridge U.P., Cambridge, MA), pp. 355–386.

Kuhl, P. K., and Miller, J. D. (**1975**). ''Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants,'' Science **190,** 69–72.

Kuhl, P. K., and Miller, J. D. (**1978**). ''Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli,'' J. Acoust. Soc. Am. **63,** 905–917.

Liberman, A. M., and Mattingly, I. G. (**1985**). ''The motor theory of speech perception revised,'' Cognition **21**, 1–36.

Lindblom, B. (**1963**). ''Spectrographic study of vowel reduction,'' J. Acoust. Soc. Am. **35,** 1773–1781.

Lotto, A. J., and Kluender, K. R. (**in press**). ''General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification,'' Percept. Psychophys.

Mann, V. A. (**1980**). ''Influence of preceding liquid on stop-consonant perception,'' Percept. Psychophys. **28,** 407–412.

Mann, V. A. (**1986**). ''Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners perception of English 'l' and 'r',''' Cognition **24**, 169–196.

Miller, J. L., and Liberman, A. M. (**1979**). ''Some effects of later-occurring information on the perception of stop consonant and semivowel,'' Percept. Psychophys. **25,** 457–465.

Ostry, D. J., Gribble, P. L., and Gracco, V. L. (**1996**). ''Coarticulation of jaw movements in speech production: Is context sensitivity in speech kinematics centrally planned?,'' J. Neurosci. **16,** 1570–1579.

Port, R. F., and Dalby, J. (**1982**). ''Consonant/vowel ratio as a cue for voicing in English,'' Percept. Psychophys. **32,** 141–152.

Raphael, L. F. (**1972**). ''Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in English,'' J. Acoust. Soc. Am. **51,** 1296–1303.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (**1994**). ''On the perceptual organization of speech,'' Psychol. Rev. **101,** 129–156.

Remez, R. E., Rubin, P. E., Carrell, D. B., and Carrell, T. D. (**1981**). ''Speech perception without traditional speech cues,'' Science **212,** 947–950.

Shepard, R. N. (**1984**). ''Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming,'' Psychol. Rev. **91,** 417–447.

Sherif, M., Taub, D., and Hovland, C. I. (**1958**). ''Assimilation and contrast effects of anchoring stimuli on judgments,'' J. Exp. Psychol. **55,** 150–155.

Ullman, S. (**1984**). ''Maximizing rigidity: the incremental recovery of 3-D structure from rigid and nonrigid motion,'' Perception **13**, 255–274.

Wallach, H. (**1948**). ''Brightness constancy and the nature of achromatic colors,'' J. Exp. Psychol. **38,** 310–324.

Warren, R. M. (**1985**). ''Criterion shift rule and perceptual homeostasis,'' Psychol. Rev. **92,** 574–584.