

TR - H - 042

**An analysis of the dimensionality of
jaw motion in speech**

**Eric Vatikiotis-Bateson
David J. Ostry**

1993. 12. 16

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 ☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1011

Facsimile: +81-7749-5-1008

AN ANALYSIS OF THE DIMENSIONALITY OF JAW MOTION IN SPEECH

Eric Vatikiotis-Bateson and David J. Ostry

*ATR Human Information Processing Research Laboratories, Japan
McGill University, Montreal, Canada*

Submitted, Journal of Phonetics
December, 1993

Abstract

The human jaw moves in three spatial dimensions, and its motion is fully specified by three orientation angles and three positions. Using OPTOTRAK, we characterize the basic motions in these six degrees of freedom and their interrelations during speech. As has been reported previously, the principle components of jaw motion fall primarily within the midsagittal plane, where the jaw rotates downward and translates forward during opening movements and follows a similar path during closing. In general, the relation between sagittal plane rotation and horizontal translation (protrusion) is linear. However, speakers display phoneme-specific differences in the slope of this relation and its position within the rotation-translation space. Furthermore, instances of pure rotation and pure translation are observed. These findings provide direct support for the claim that jaw rotation and translation are independently controlled (Flanagan, Ostry, & Feldman, 1990). Rotations out of the midsagittal plane are also observed. Yaw about the longitudinal body axis is approximately 3 degrees and roll usually less than 2 degrees. The remaining non-sagittal component, lateral translation, is small in magnitude and uncorrelated with other motions.

Introduction

The human jaw is a rigid skeletal structure whose position and orientation in space are fully described by three orientation angles and three positions (Figure 1). The jaw may translate along vertical, horizontal and lateral axes and may rotate about each axis as well. In speech, jaw motion has typically been studied only in the midsagittal plane. In this plane, jaw motion involves a combination of rotation about a transverse axis through the condyles and a combination of vertical and horizontal translation (Edwards & Harris, 1990; Ostry & Munhall, in press; Westbury, 1988). The jaw rotates downward and translates forward during opening; during closing, the pattern is reversed.

Jaw motions are produced by muscles which have multiple mechanical actions. Consequently, there is no one-to-one relation between muscle actions and kinematic degrees of freedom. However, despite this complex relation, control of the jaw's motion in the sagittal plane appears to coincide with the jaw's mechanical degrees of freedom (Flanagan, et al., 1990; Ostry & Munhall, in press). Specifically, when loud and normal volume speech were compared by plotting jaw rotation as a function of horizontal jaw translation, rotation and horizontal translation varied independently. This suggests that jaw rotation and translation are separately controlled and that muscle actions must be coordinated in such a way as to produce motion in individual degrees of freedom as well as in combination.

In this paper, we mathematically decompose three-dimensional jaw motions during speech into the six component orientations and positions and examine their variation across a variety of phonetic and speaking conditions in order to address several issues. First, what are the major kinematic components of jaw motion during speech, and how do they interact? For example, to what extent can speech articulation be satisfactorily described within the midsagittal plane (Stone, 1990)? Moreover, are all midsagittal components necessary to the description? Second, what is the relation between the mechanical degrees of freedom, which result from the rigid body decomposition, and the underlying control of 3D jaw motion? For example, can the control of 3D jaw motion be accounted for in terms of separate commands for jaw rotation and translation, as has been hypothesized for 2D motion (Flanagan, *et al.*, 1990).

Methods

Subjects and stimuli

Four native speakers of English (2) and Japanese (2) produced repetitive sequences of symmetrical VCVCa utterances, such as *asasa* and *arara*, in normal rate, fast rate, and

loud speaking conditions. The consonants were /s, ʃ, f, p, t, k, r/ (and /l/ for English), and the vowels, /i, a, e, o/. In Japanese, *isisa* and *ifi/a* are homophonous and there is no phonemic /l/; therefore, Japanese speakers produced only 27 (of 32) utterance sequences per condition. Each utterance sequence consisted of at least 10 repetitions. Language-specific differences in accentual patterning were preserved in the nonsense utterances; Japanese utterances were slightly more prominent on the first syllable, while English utterances had a stressed second syllable.

Equipment and data recording

OPTOTRAK (Northern Digital, Inc.) was used to record the three-dimensional positions of 12 infra-red (IR) markers attached to the head (6) and jaw (6). Markers were attached to a block of styrofoam mounted on a headband and to a steel and acrylic jaw splint. System accuracy was between 0.003 mm (static) and 0.05 mm (dynamic). Marker positions were sampled at 200 Hz. The raw position data were low-pass filtered at 10 Hz with a bi-directional second order Butterworth filter. This filter frequency corresponded to a signal power approximately 60dB below peak signal power.

Coordinate transformation and rigid body reconstruction

Static trials and measures of the distance from the jaw condyle to the lower front incisors were used along with vendor-supplied software to perform two coordinate transformations needed to define rigid body jaw motion for each speaker. The first transform removed head motion and was used to generate head-corrected 3D position data. The second coordinate transform was used to decompose the jaw's motion into constituent rotation angles and translation positions for each axis. The derived frame of reference of the jaw is shown schematically in Figure 1. Axes through the jaw condyles and the occlusal and midsagittal planes, define the coordinate system. Translation along each axis is referred to by the name of the axis and, per convention, the three rotations are roll, pitch, and yaw about the horizontal, lateral, and vertical axes, respectively. The two-stage rigid body transform employs a quaternion method of rigid body reconstruction (Horn, 1987) and iterative regression estimation of 6D orientation values on a hardware-specific model of the known marker positions. The calculated error of the first transform, which generated head-corrected 3D positions, was negligible; error for the second transform was less than 5% of a unit of rotation (degree) or translation (mm).

Figure 1
-----*Data analysis*

Tangential velocities and accelerations were derived from the 3D positions of the jaw marker closest to the front incisors using a central difference algorithm. Velocity peaks and acceleration zero-crossings were used to identify onsets and offsets of the consonant-to-vowel transitions (jaw opening) for the first CV and vowel-to-consonant transitions (jaw closing) for the second VC. From these onsets and offsets, jaw orientation angles, positions, and motion paths were derived. This scoring method worked well under most conditions. However, when trajectory amplitude was very small, the derivatives of these movements were often too noisy for the algorithm to work successfully. Since hand measurement of these cases was usually found to be unreliable as well, they were discarded. This affected the corpus most for /i/ context utterances, particularly at the fast speaking rate.

Results*Defining the major components of motion*

Kinematic studies of speech articulation have typically dealt with only one or at most two dimensions of motion (Edwards, 1985; Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985; Kuehn & Moll, 1976; Ostry, Keller, & Parush, 1983). Recently, researchers have tried to assess the three-dimensional behavior of articulator structures by combining separately obtained two-dimensional data using various imaging and position sensing techniques (Stone, 1990; Stone & Vatikiotis-Bateson, submitted). By and large, simultaneous transduction of 3D motion has been outside the purview of speech research due to the combination of technical limitations and the operating assumption that the relevant aspects of speech behavior are recoverable from the midsagittal plane. In this paper, we examine 3D motion in two ways. First, we provide a brief description of the 3D motion of a jaw marker near the incisors. Second, we describe the motion of the jaw as a whole; that is, as a rigid body whose motion has six component degrees of freedom.

Figure 2

3D motion

Figure 2 shows the three dimensions of motion of a jaw marker as a function of time for several loud repetitions of *asasa*, produced by one of the English speakers (EVB). The data have been corrected for head movement and are expressed as distances from the coordinate system origin shown in Figure 1. The figure shows the movement of the the jaw splint marker nearest the teeth. This marker is approximately 4 cm in front of the lower incisors and lies almost on the midsagittal plane. The movement amplitudes are larger than they would be if transduced at the lower teeth (see below); however, spatiotemporal patterning is not affected by marker position.

All three traces are correlated with one another and with the phonetic events in the production. Peaks in the horizontal (towards the camera) and vertical axes correspond to closures for the consonant, /s/, and valleys correspond to the vowel, /a/. As the jaw opens for /sa/, the marker moves down and slightly back (away from the camera). Lateral motion of this marker for these utterances is further from midline during the vowel than the consonant (see position scale). Overall, lateral deviations from the midsagittal plane were found to be small for all speakers and utterance conditions, ± 2 mm on average and never more than ± 7 -8mm ((Vatikiotis-Bateson, Gribble, & Ostry, 1993); for details, see (Vatikiotis-Bateson, Gribble, & Ostry, submitted)).

6D rigid body reconstruction

There is an important difference between kinematic analysis of motion typical in speech and rigid body reconstruction of that motion: In 3D analysis of motion, measurements made at different points on an object such as the jaw will usually result in different trajectories in 3-space. A good example is the one just given above in which the motion of a marker protruding from the jaw on a splint is larger than motion at a point on the front teeth. By contrast, in rigid body analysis, every point on the object moves the same way. That is, the motion of the entire object is considered rather than the motion of a point. Using motion of markers on a splint rigidly attached to the jaw, we can describe motion of the jaw itself.

 Figure 3

Figure 3 shows the six reconstructed orientation angles and positions as a function of time for the *asasa* productions shown in Figure 2. These data are typical of the entire set in which five of the six components exhibited smooth and correlated patterning through time (lateral motion was very noisy and, at best, only weakly correlated with the others). Peaks and valleys of these five time series generally coincided with consonant closure and vowel opening. The largest components of the motion were the three acting within the mid-sagittal plane; namely, pitch rotation about a transverse axis through the temporomandibular joint (TMJ) and the translations along the vertical and horizontal axes. In this example, the jaw rotated downward and back through an arc of about 15 degrees and translated horizontally forward (protrusion) approximately 10 mm and vertically downward 3-4 mm. Roll about the horizontal and yaw about the vertical axes were much smaller, accounting for about 1 and 3 degrees of rotation, respectively. Lateral translation was less than 1 mm and quite noisy. Thus, as the speaker opened his jaw for production of /sa/, the jaw moved mainly in the mid-sagittal plane.

Midsagittal components

Since the analyses of Edwards (1985; Edwards & Harris, 1990) and Westbury (1988), it is now generally accepted in speech that midsagittal jaw motion entails both rotation about the TMJ and some combination of horizontal and vertical translation (Flanagan *et al.*, 1990; Ostry & Munhall, in press). In this section, we examine the relations among midsagittal components by plotting the motion of one against the other.

Pitch rotation vs. horizontal translation

In Figure 4a, pitch rotation is plotted against horizontal translation throughout the course of opening CV gestures for loud productions of *asasa* and *arara*. The nearly straight line paths shown in the figure indicate a nearly constant relation between pitch rotation and horizontal translation. This systematic relation is consistent with previous studies (e.g., Flanagan *et al.*, 1990; Ostry & Munhall, in press). However, as also seen previously, the slope of the rotation-translation relation depended on phoneme context. For the loud productions of *asasa* and *arara* shown in Figure 4a, the jaw arrived at almost the same orientations and positions for the vowel /a/ from quite different starting positions

for /s/ and /r/. Similar phoneme-specific patterns were observed for all three speaking conditions.

 Figure 4(a-c)

Indeed, every speaker showed some degree of consistent phoneme-specific patterning in the relation of pitch rotation and horizontal translation across changes of speaking condition, but the extent and nature of the interaction of speaking condition with phoneme context was speaker dependent. Figures 4b and 4c show opening (/sa/) trajectories during normal, loud, and fast productions of *asasa* for the two Japanese speakers. Comparing the loud productions of these two speakers with those of the English speaker shown in Figure 4a, it is clear that the magnitudes of rotation and especially translation were much smaller for the Japanese speakers — 6 vs. 15 degrees of rotation and less than 3 vs. 10 mm of translation.

Despite the substantial magnitude difference between the Japanese and English speakers (rotation and translation values for the second English speaker, DJO, were only slightly smaller than those of speaker EVB), all speakers showed progressive reduction in magnitudes of rotation and translation from loud to normal to fast speaking conditions. Scaling of movement amplitude with speech rate and/or volume has been seen in almost all studies of articulator motion (cf. (Gay, 1981), who showed rate distinctions *can be* produced without changing movement amplitude). Other effects of speaking condition were more idiosyncratic. For example, as shown in Figures 4b and 4c for *asasa*, speaking condition affected the slope of the rotation-translation relation and the relative position of the trajectory quite differently for the two Japanese speakers. Slopes of speaker MH's trajectories were progressively steeper from loud to fast, and initial positions for jaw translation (at consonant closure) were progressively protruded. MH's data look as if the paths were converging on a target vowel configuration of rotation and translation. At the faster rate, movements consisted of almost pure rotation (translation was often less than 1 mm). For speaker YHK, on the other hand, there was little difference in slope of the rotation-translation relation, and initial translation positions for both fast and loud productions were more retracted (smaller) than those of normal rate productions.

 Figure 5

Consonant- and vowel- specific differences in movement paths were realized as differences in the slope, intercept, and/or starting points of the rotation-translation trajectory. This is exemplified for English speaker EVB in Figure 5, which shows trajectories for all consonants in the /i/ context. Trajectories for the three alveolars, /s,sh,t/ overlapped, while /k/ trajectories had the same slopes but the jaw began to move at the position and orientation angle where the alveolar trajectories ended for the vowel. Bilabial trajectories, on the other hand, also had slopes similar to the alveolars and /k/, but their initial horizontal positions, hence their intercepts on the horizontal position axis, were shifted backwards (more retracted). Similarly, trajectories for /l, r/ were retracted but differed from the bilabials in slope and initial orientation. The case of /l/ is particularly interesting, because it demonstrates almost pure jaw translation (as opposed to the almost purely rotational motions produced in some contexts by the Japanese speakers).

Also, some contexts resulted in more variable trajectories than others. For example, /a/ was the most stable context for all speakers, while /r/ and /p/ were the most variable, especially for the Japanese speakers.

Finally, although the relation between pitch rotation and translation was usually a smooth, quasi-linear function as shown in Figures 4 and 5, there were many instances of substantially greater curvature in the function, especially in contexts where the anterior tongue was not raised — e.g., /p,f,r/ followed by /o/ for Japanese speaker YHK.

Vertical vs. horizontal translation

The third midsagittal component, vertical translation, was highly correlated with the other two in the data of the English speakers, DJO and EVB (see Fig. 3). Figure 6a plots vertical against horizontal translation for all loud productions of speaker EVB — i.e., 8 consonant and 4 vowel contexts. As shown, the amount of vertical translation downward during opening gestures was about one quarter the amount of horizontal protrusion, and the function described a fairly smooth curve. Figure 6b shows data for Japanese speaker YHK. The curvature of the distribution is clearly similar to that seen for EVB's data, but is composed of short irregular trajectories characteristic of YHK's small horizontal translations.

Figure 6(a-c)

How is this curvature achieved? Unlike pitch rotation and horizontal translation — components whose contribution to motion varies with context and therefore could be independently controlled parameters — all of the trajectories (for all speakers) fall along a similar curve. This suggests a structural rather than functional constraint. We suspect vertical translation is a consequence of the anatomy of the TMJ. Figure 6c (from (McDevitt, 1989)) shows an anatomical cutaway of the lateral pterygoid muscle and TMJ. The curved shape of the articular eminence, along which the jaw condyle moves, is clearly quite similar to that of the trajectories shown for the two speakers. This would account for the quasi-linear relation between the translation components shown in Figure 6a. However, when jaw motion is primarily rotational and horizontal translation is small (Fig. 6b), we suggest that the curved shape is determined by the different regions along the articular eminence — i.e., different horizontal positions — along which jaw rotation occurs.

Non-sagittal components

The two non-sagittal rotations — yaw about the vertical axis and roll about the horizontal axis — also demonstrated systematic patterning across speakers and conditions (see Figs. 2 and 3). However, as already mentioned, their combined effect on lateral 3D motion at the incisors was small and, in our opinion, does not seriously distort or invalidate 2D measurements made with midsagittally restricted devices such as the x-ray microbeam or electromagnetometer. As shown for EVB's data in Figure 3, yaw was typically 3 degrees or less and roll 2 degrees or less. Magnitudes of yaw and roll angles differed little across speaking conditions, though a small constant increase in roll was often observed for the loud speaking condition. Analysis of the phoneme-specific effects has not revealed any systematic patterning of the non-sagittal components.

An interesting finding concerning non-sagittal motion stemmed from the question (posed by Kevin Munhall) of whether the orientation of the frame of reference changes the data significantly. This is an important issue because bias introduced by the seemingly arbitrary, though conventional, choice of the midsagittal, occlusal plane orientation could affect interpretation of the results. For example, by changing the

orientation of the reference frame within the midsagittal plane from the occlusal bite plane to a plane passing through the articular eminence, the roll component was effectively eliminated as was the monotonic covariation between horizontal and vertical translation (Vatikiotis-Bateson *et al.*, 1993). Thus, choice of orientation affects the results and cannot be arbitrary (for detailed discussion, see Vatikiotis-Bateson *et al.*, submitted). However, a change of orientation, in the present data set, from the occlusal plane to the plane of the articular eminence did not substantially alter the pattern of the relation between pitch rotation and horizontal translation.

Discussion

In summary, we have shown that when 3D motion of the jaw is decomposed into the three rotations and three translations which fully characterize its motion, all components except lateral translation may be correlated with one another over the course of the movement. Since the anatomy of the TMJ allows very little lateral translation, the small amount of lateral motion observed at the teeth was due to yaw about the vertical axis and perhaps to roll (depending on the coordinate system orientation).

The principal components of jaw motion during speech lie within the midsagittal plane. When pitch translation is plotted against horizontal translation, nearly linear paths occur. Furthermore, the slopes and intercepts differ according to the consonant-vowel composition of the utterance. Instances of pure rotation and pure translation were observed in addition to the more typical combination of the two.

Vertical and horizontal translation were also correlated. However, there was no context-specific patterning and the form of the movement paths relating vertical and horizontal components was similar to that of the articular eminence of the upper skull along which the condyle moves. Thus, we propose that the form of the vertical translation arises from the anatomical structure of the TMJ and is not directly controlled.

These findings provide direct support for the idea that the control of jaw motion in speech involves the independent specification of sagittal plane rotation and horizontal translation. Specifically, the data demonstrate that specific utterances may be achieved by rotation alone and translation alone. Independent control of rotation and translation is a basic notion associated with the model for jaw movement proposed by Flanagan *et al.* (1990). The model proposes that the observed straight-line paths for pitch rotation against horizontal translation arise when the independently specified equilibrium position and orientation are shifted simultaneously and with the same relative velocity.

While the slopes and intercepts of rotation on translation varied for different consonant-vowel combinations, we saw no evidence of phoneme-specific targets as would be indicated by converging paths for specific vowels or consonants. During speech, control of jaw motion must be coordinated with that of other articulator structures, such as the tongue and lips, which more directly effect specific vocal tract configurations. Thus, phoneme-specific articulatory targets, if they occur at all, need not be specified at the level of the jaw. Nevertheless, the global specification of speech goals, either at vocal tract or acoustic levels, is apparently organized to permit jaw control resulting in straight line paths.

Acknowledgment

The authors thank Paul Gribble, Shinji Maeda, and Yoh'ichi Tohkura for their assistance and support.

References

- Edwards, J. (1985) *Mandibular rotation and translation during speech*. Unpublished Doctoral Dissertation, CUNY.
- Edwards, J., & Harris, K. S. (1990). Rotation and translation of the jaw during speech. *Journal of Speech and Hearing Research*, **33**, 550-562.
- Flanagan, J. R., Ostry, D. J., & Feldman, A. G. (1990). Control of human jaw and multi-joint arm movements. In G. E. Hammond (Eds.), *Cerebral control of speech and limb movements*. Amsterdam: Elsevier Science Publishers (North-Holland).
- Gay, T. J. (1981). Mechanisms in the control of speech rate. *Phonetica*, **38**, 148-158.
- Horn, B. K. P. (1987). Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, **4**, 629-642.
- Kelso, Vatikiotis-Bateson, E., Saltzman, E. L., & Kay, B. A. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America*, **77**, 266-280.
- Kuehn, D. P., & Moll, K. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, **4**, 303-320.
- McDevitt, W. E. (1989). *Functional anatomy of the masticatory system*. London: Wright.
- Ostry, D. J., Keller, E., & Parush, A. (1983). Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. *Journal of Experimental Psychology: Human Perception and Performance*, **9**, 622-636.

- Ostry, D. J., & Munhall, K. G. (in press). Control of jaw orientation and position in mastication and speech. *Journal of Neurophysiology*.
- Stone, M. (1990). A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data. *Journal of the Acoustical Society of America*, **87**, 2207-2217.
- Stone, M., & Vatikiotis-Bateson, E. (submitted). *Journal of Phonetics*.
- Vatikiotis-Bateson, E., Gribble, P., & Ostry, D. J. (1993). Functionality of jaw motion components during speech. *Acoustical Society of Japan*, **5**(10), 277-278.
- Vatikiotis-Bateson, E., Gribble, P., & Ostry, D. J. (submitted). Functional geometry of jaw motion in speech. *Journal of the Acoustical Society of America*.
- Westbury, J. R. (1988). Mandible and hyoid bone movements during speech. *Journal of Speech and Hearing Research*, **31**, 405-416.

Figure Captions

Figure 1. Frame of reference for jaw motion showing the three coordinate axes for translation and rotation. The origin is defined with the horizontal axis aligned with the occlusal plane.

Figure 2. 3D position of the jaw marker nearest to the teeth plotted over time for loud productions of *asasa* by English speaker EVB. Position scales for the smaller horizontal and lateral motions have been expanded relative to the vertical (top).

Figure 3. Jaw rotations (upper panels) and translations (lower panels) during loud volume repetitions of *asasa*, produced by speaker EVB. Pitch rotation and horizontal translation are the largest amplitude components. Lateral translation is small and uncorrelated with the other motions.

Figure 4(a-c). a. Pitch rotation versus horizontal translation for opening gestures during loud productions of *asasa* (solid) and *arara* (dashed), produced by speaker EVB. b-c. Fast (dotted), normal (dashed), and loud (solid) speaking conditions for two Japanese speakers, MH (b) and YHK (c). Fast movements for both Japanese speakers show almost no translation.

Figure 5. Pitch rotation versus horizontal translation during loud productions of all consonants in the /i/ context, produced by speaker EVB. Repetitions of *ilila* (solid) involve almost pure translation.

Figure 6(a-c). a-b. Vertical versus horizontal jaw translation in loud speech: a. English speaker EVB for all consonants and vowels; . b. Japanese speaker YHK for all consonants and vowels (no /i/). c. The overall shape of these motion paths corresponds to the shape of the articular eminence of the upper skull.

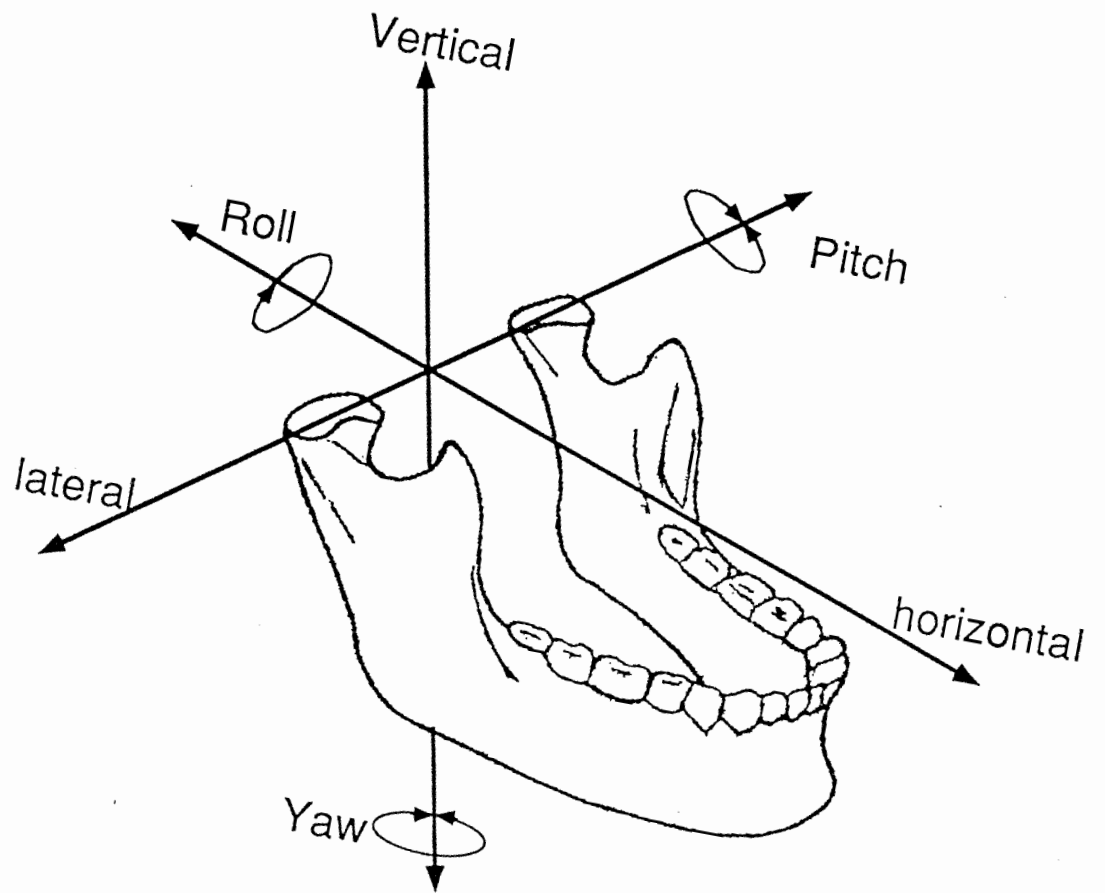


Figure 1

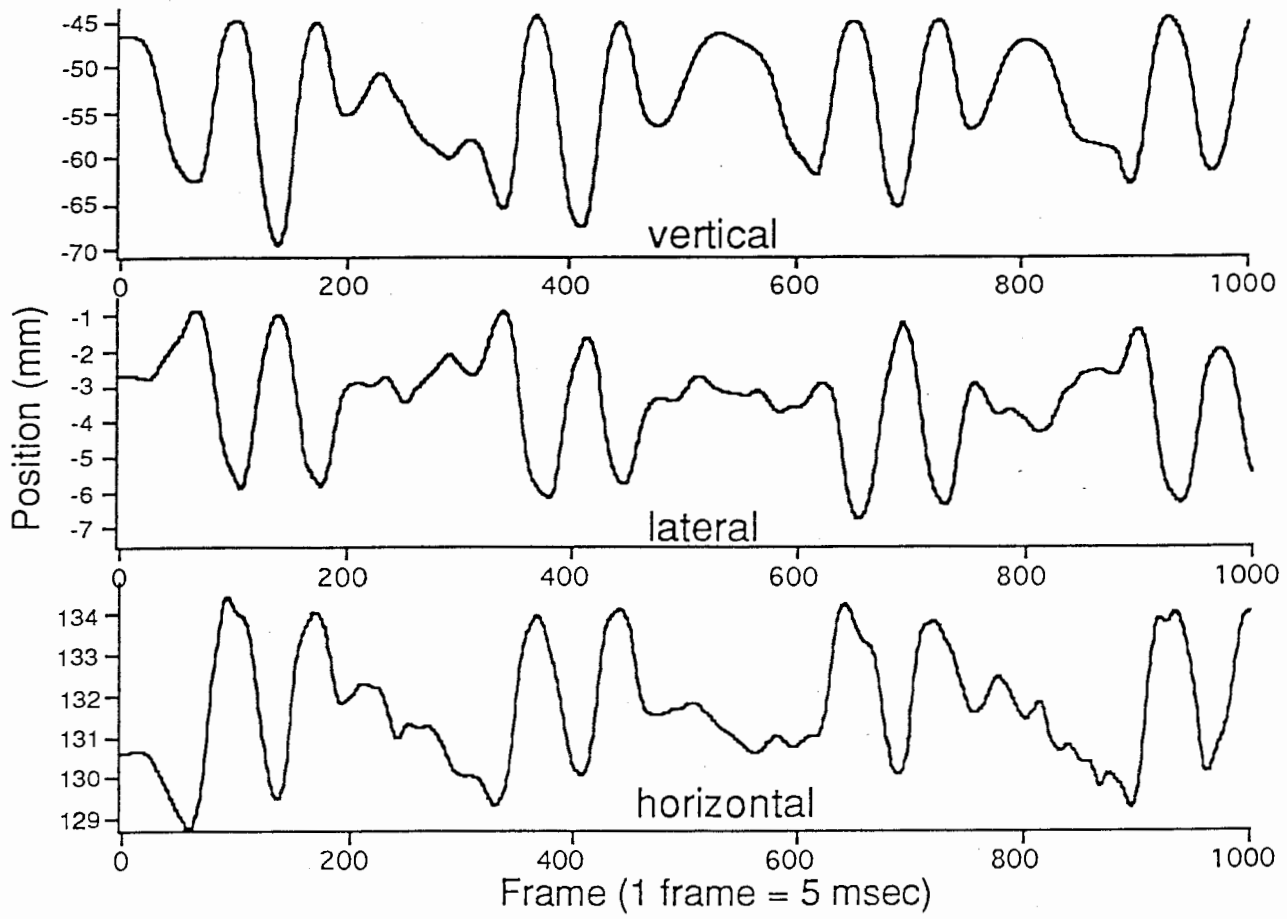


Figure 2

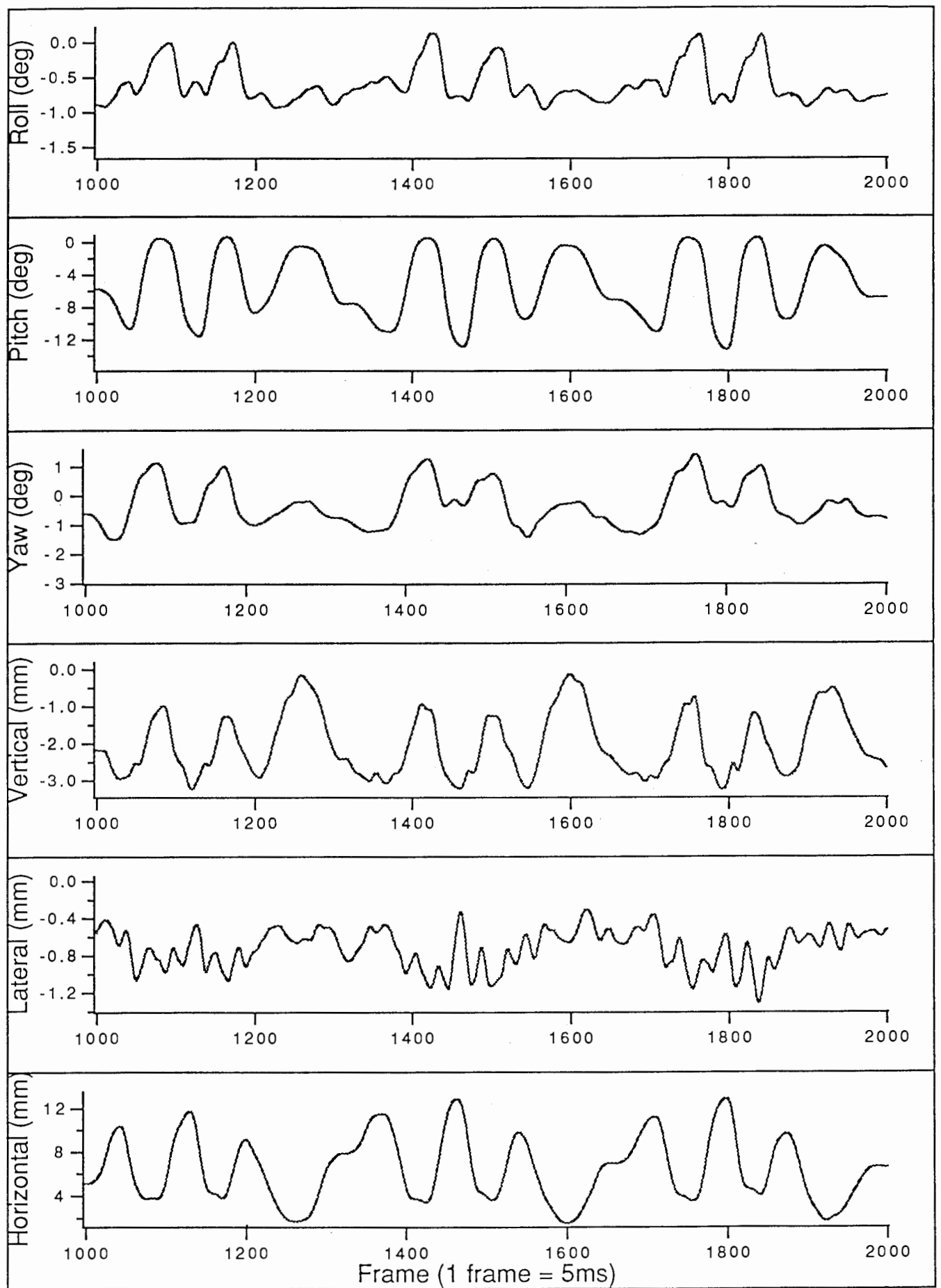


Figure 3

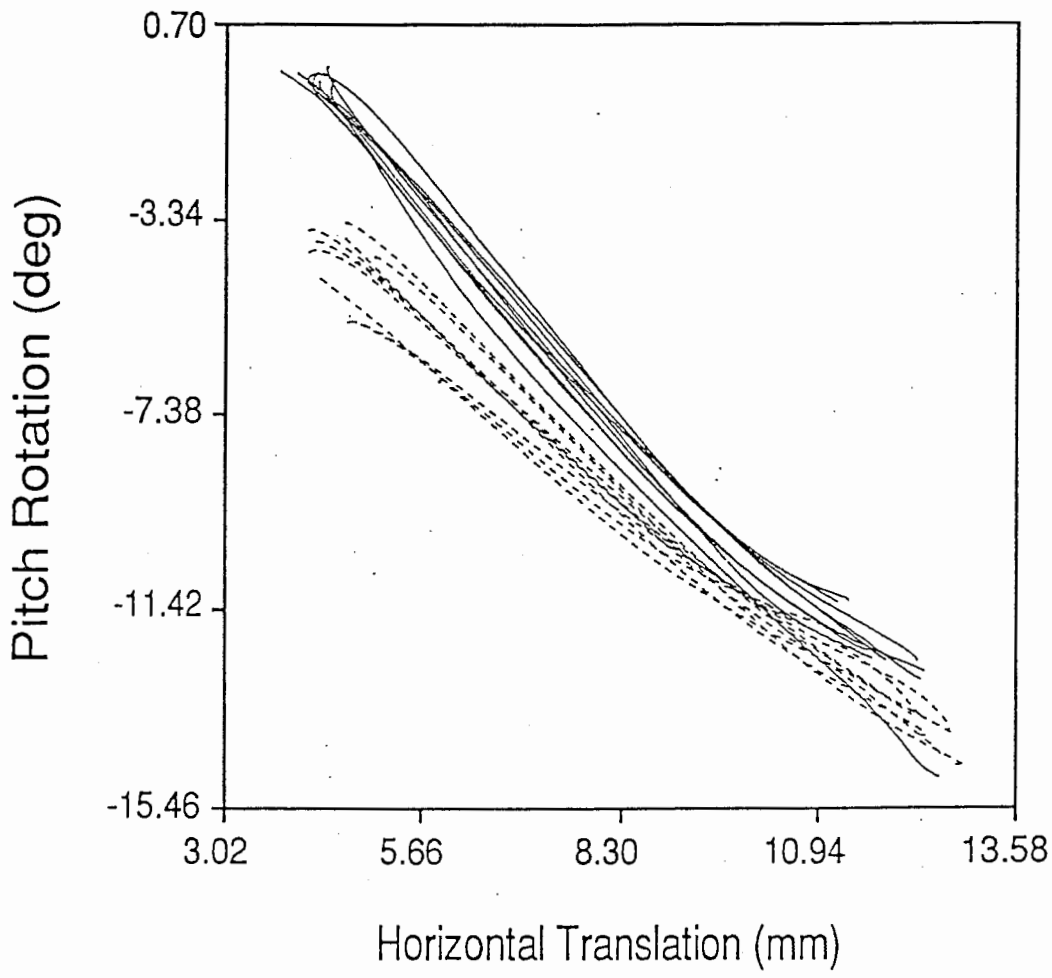


Figure 4a

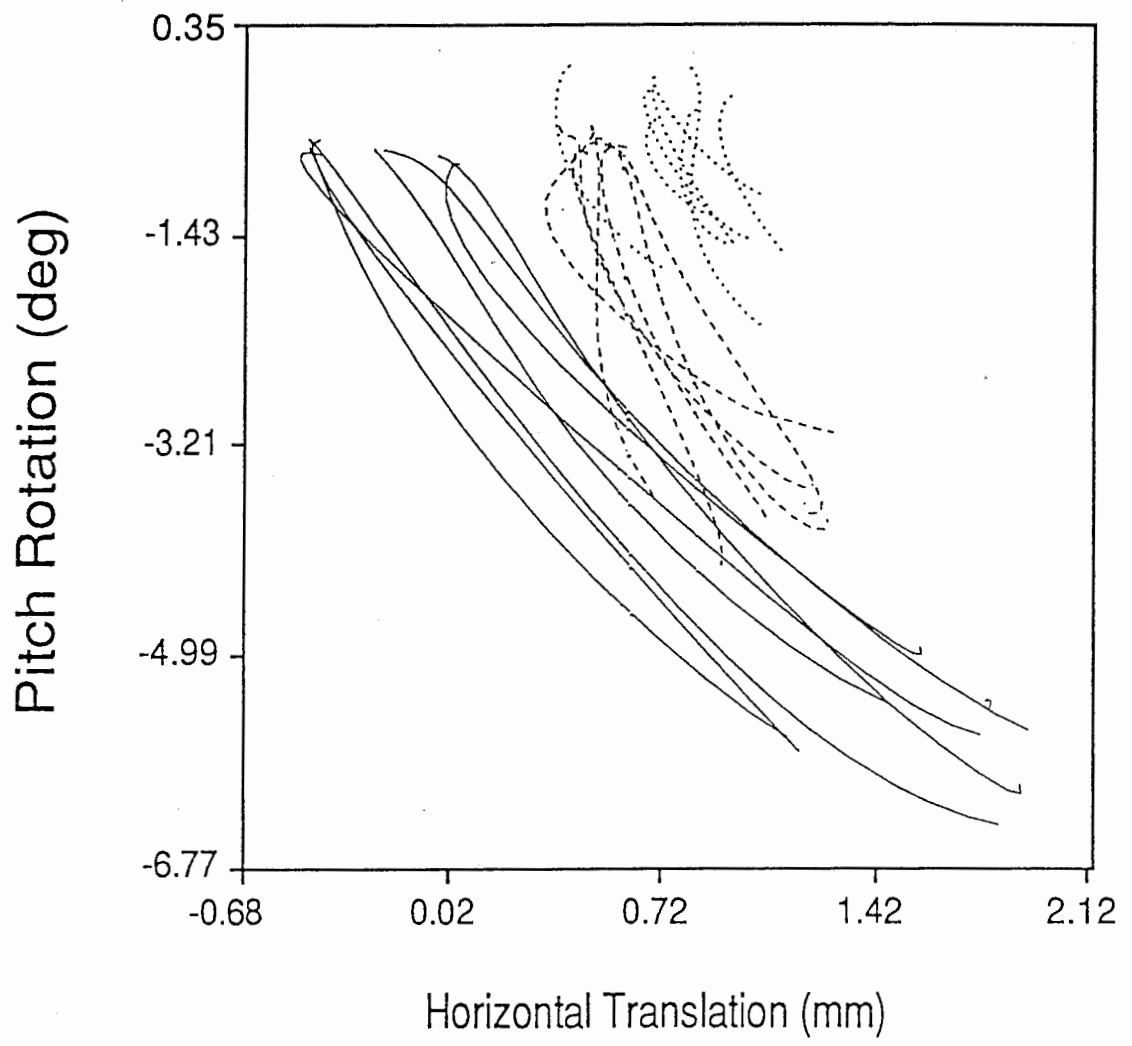


Figure 4b

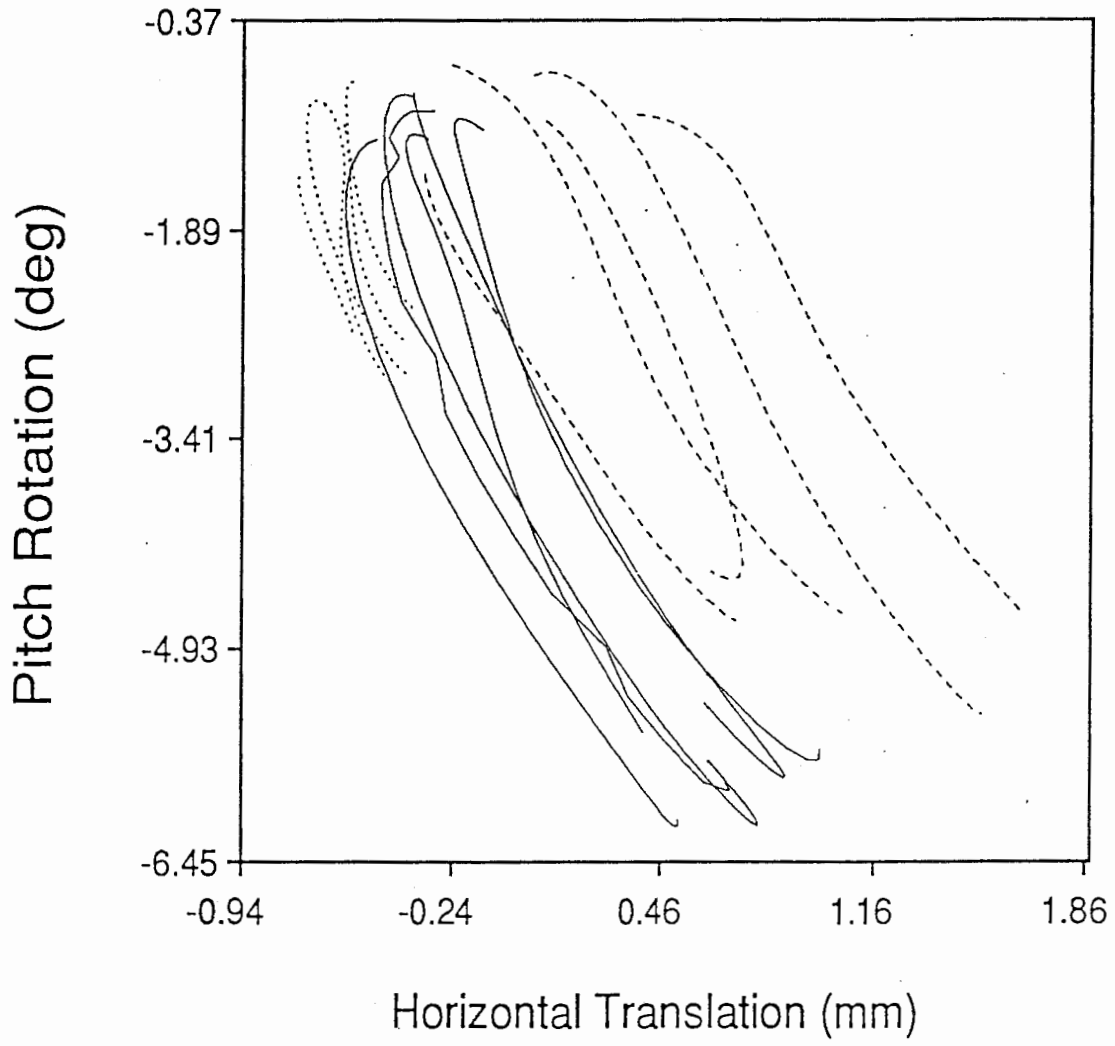


Figure 4c

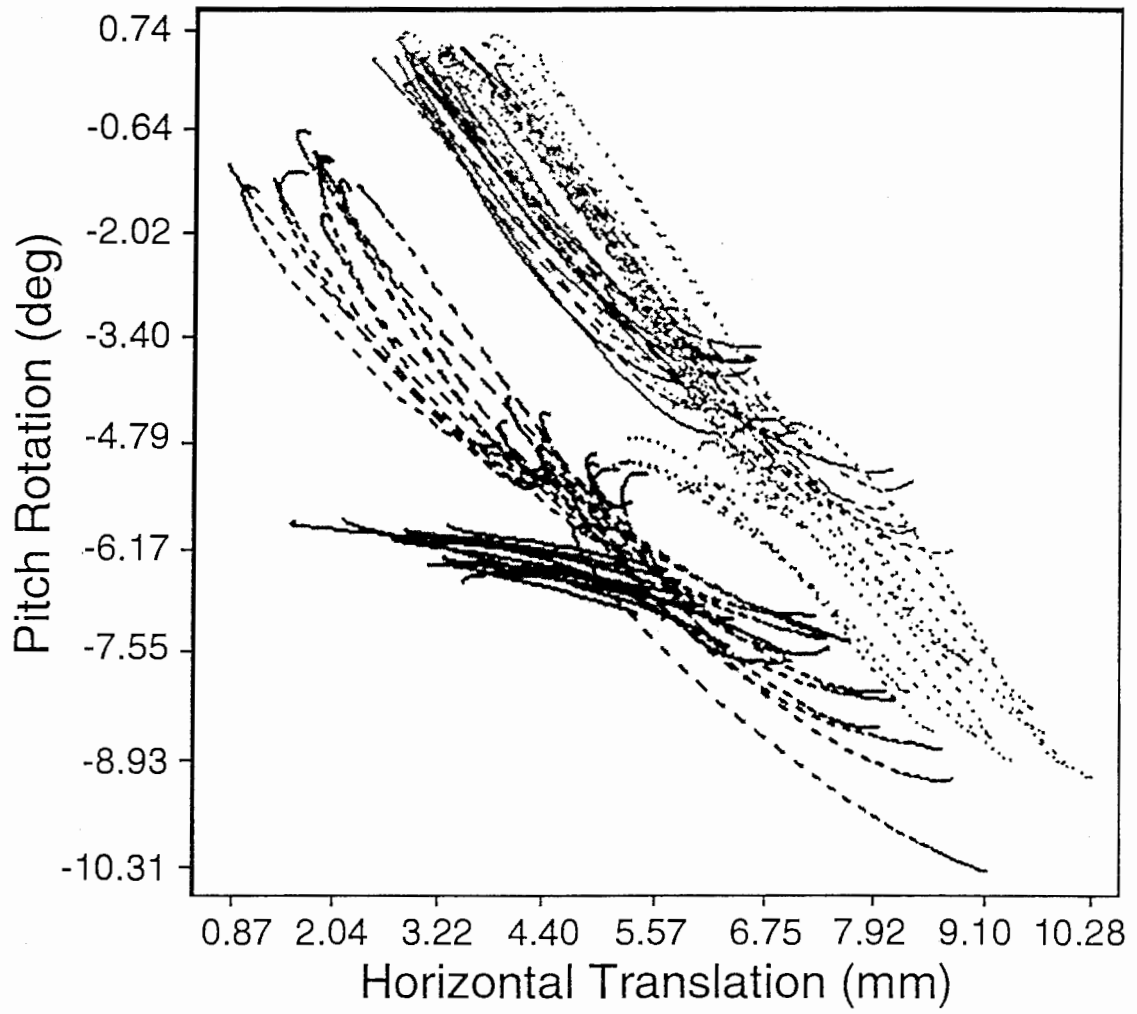


Figure 5

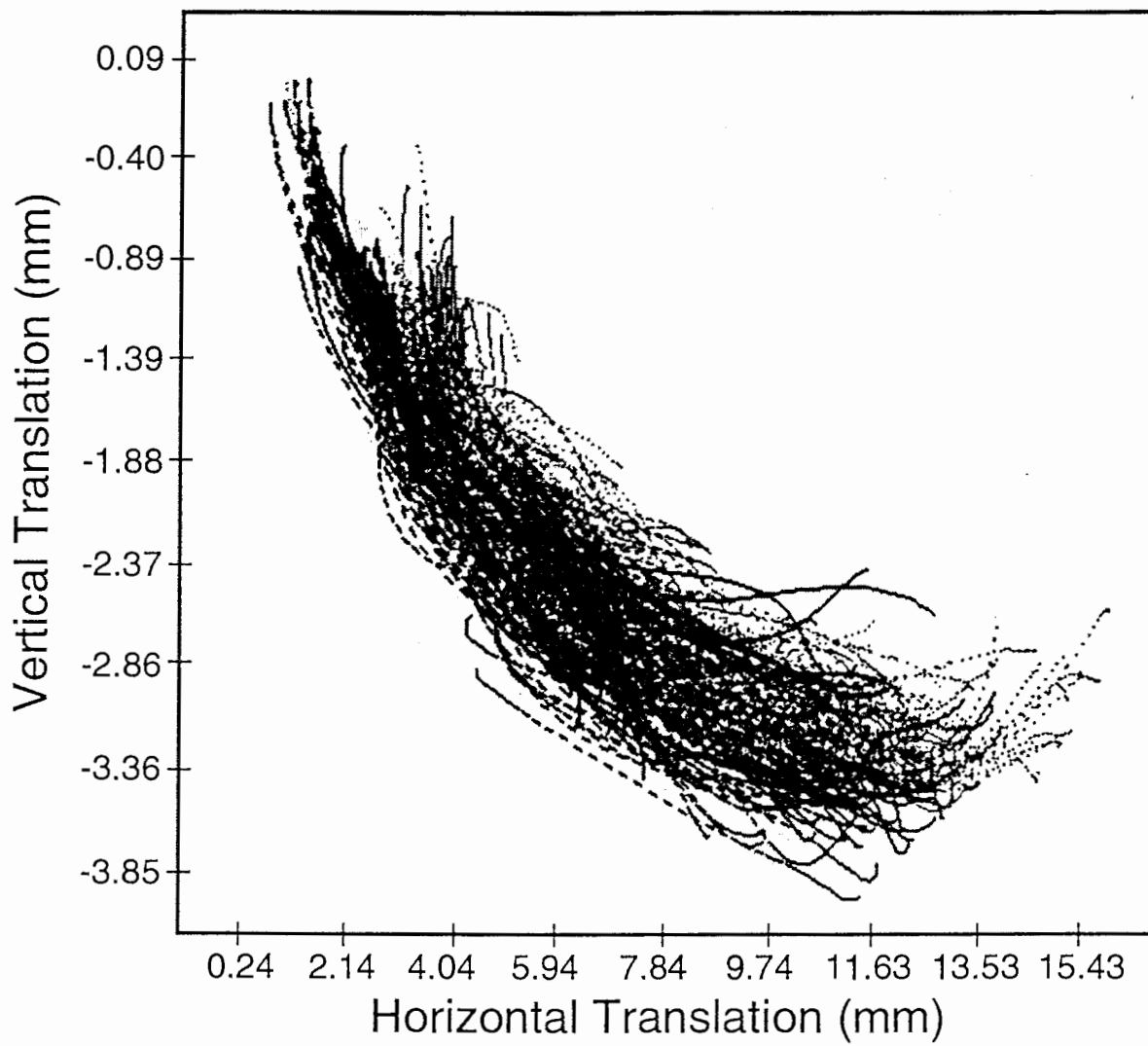


Figure 6a

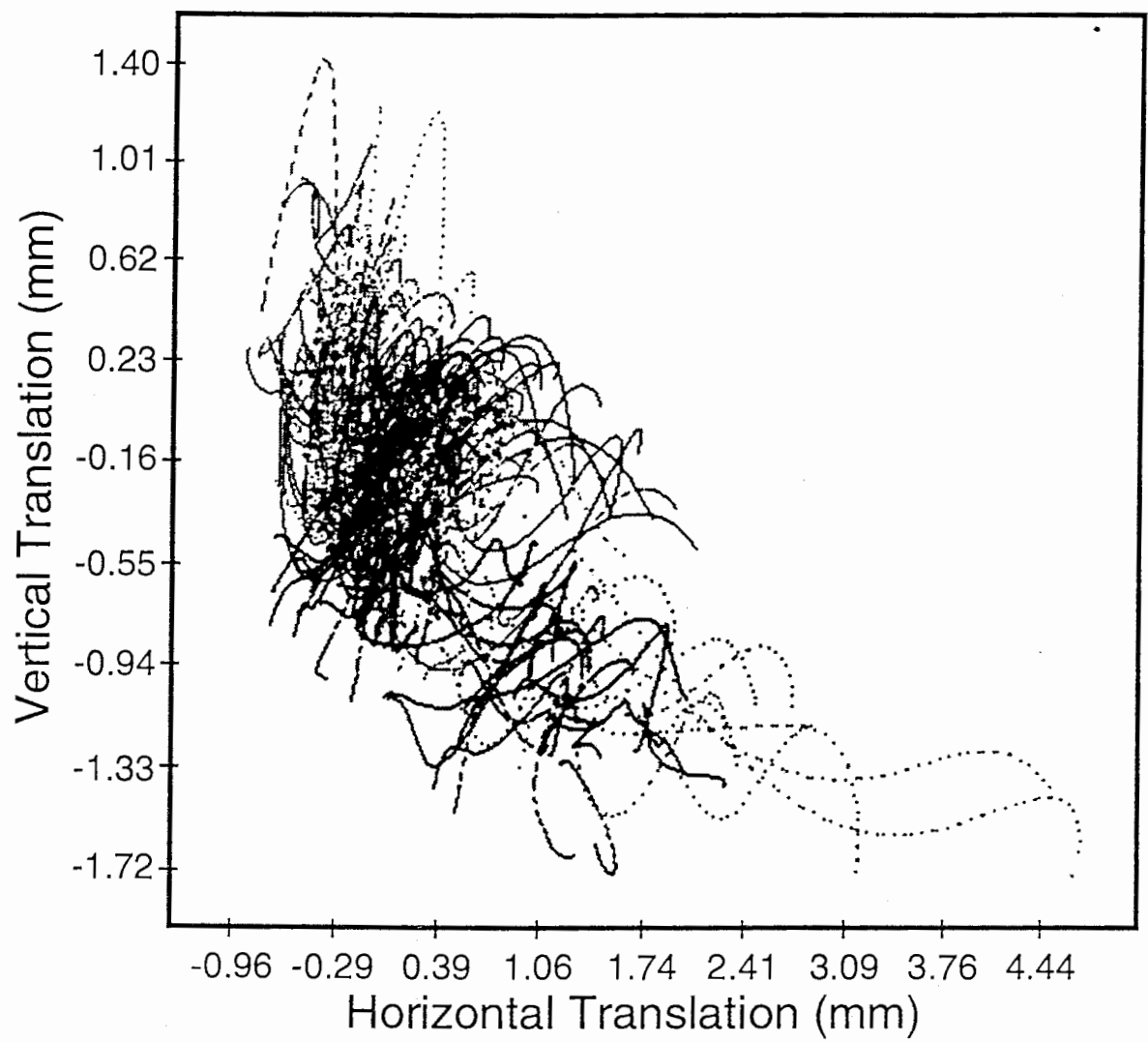


Figure 6b

Articular eminence

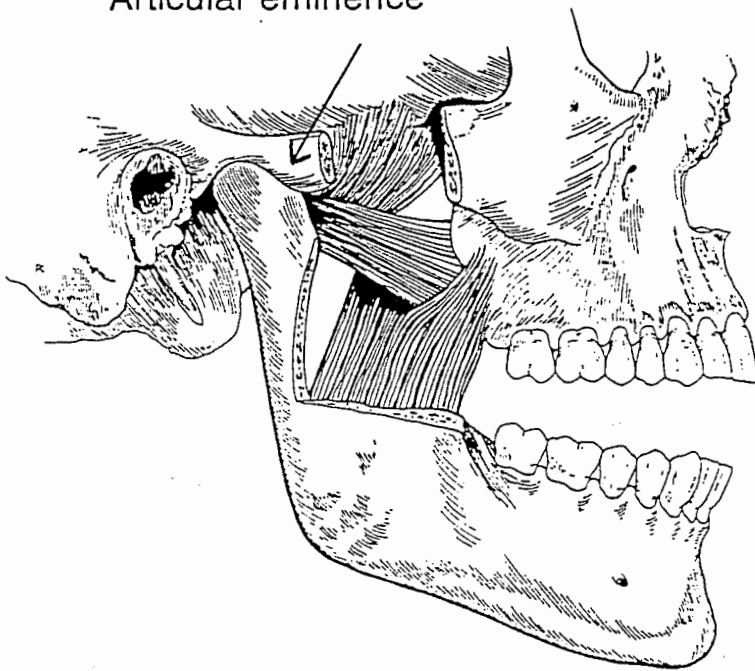


Figure 6c