Role of respiratory control mechanisms in the pathogenesis of obstructive sleep disorders

Magdy Younes
Sleep Centre, Foothills Medical Centre, University of Calgary, Calgary, Alberta, Canada

Obstructive ventilatory sleep disorders result when the pharynx collapses sufficiently during sleep to present a significant mechanical constraint to ventilation. The abnormalities may take the form of steady snoring, protracted hypopneas with intermittent arousals [the upper airway resistance syndrome (34)], or self-perpetuating transient obstructive events that recur over minutes, hours, or the entire sleep time. Frequently, these various manifestations occur in the same individual at different times (127).

Obstructive phenomena are very common. For example, habitual snoring with intermittent hypopnea/apneus occurs in 52% of women and 64% of men (134), while 4% of women and 9% of men develop >15 discrete obstructive events per hour (134). Not all affected individuals suffer clinical consequences, but there is convincing evidence that these disorders adversely affect quality of life and cognitive function in many patients (85). As well, there is growing evidence that the risk for hypertension and its complications is increased in affected individuals (31, 85).

The pharynx of patients with obstructive sleep disorders is, on average, smaller and more collapsible than in others (26, 40, 45, 94, 95). During wakefulness, this is compensated for through appropriate increases in dilator muscle activity (71, 107). In fact, these patients have no breathing difficulty while awake. For example, obstructive sleep apnea (OSA) patients achieve normal maximum exercise levels (49, 63, 64) even though this requires flow rates >4 l/s (cf. <0.5 l/s during sleep). The fundamental reason for development of obstructive sleep disorders is that the increase in dilator activity observed in wakefulness is substantially lost at sleep onset (27, 72). In subjects with more favorable pharyngeal mechanics, dilator activity also decreases at sleep onset (27, 110, 124), but airway size does not decrease enough to interfere with flow requirements during sleep. Thus these disorders develop when the normal reduction in dilator activity at sleep onset occurs in a subject whose pharynx requires relatively high dilator activity to remain sufficiently open.

A critically important observation is that the polysomnographic features vary considerably among patients with the same degree of pharyngeal collapsibility (45, 127) and even vary in the same patient at different times during sleep (127). These observations indicate that how an individual responds to the sleep-induced obstruction determines to a large extent the type and severity of the disorder (127). This review will be concerned primarily with the mechanisms that determine the response to sleep-induced obstructions, and hence the polysomnographic manifestations of the disorder.
MECHANICAL CHARACTERISTICS OF THE PASSIVE PHARYNX

An understanding of the mechanical properties of the passive human pharynx is essential to this topic because it is these passive properties that determine whether, and by how much, pharyngeal dilators must be activated to maintain adequate ventilation. Except at its extreme upper and lower ends, where it is anchored to bone and cartilage (larynx), respectively, the human pharynx has no rigid support. Accordingly, throughout most of its length its cross-sectional area varies with lumen pressure (44, 45). The pressure-area relation of the passive human pharynx has been studied in detail (44, 45). In all humans a certain lumen pressure can be found at which the passive pharynx is completely closed (P_{Close}) (44, 45). Above P_{Close}, the relation between area and pressure is nonlinear such that pharyngeal compliance (Δarea/ΔP_{LUMEN}; ΔA/ΔP) decreases as pressure and area increase (44, 45; Fig. 1A). P_{Close} and the shape of the pressure-area relation vary considerably among individuals, with the four curves shown in Fig. 1A representing most of the spectrum (45). As can be seen, at atmospheric airway pressure (P_{AW} = 0) passive pharyngeal area can be large (top curve) or small (middle curves), or the pharynx may be completely closed. A number of fixed skeletal and soft tissue features can be found at which the differences in passive mechanical behavior. Furthermore, the passive behavior of the pharynx may change in the same individual in response to changes in body and neck position, lung volume (caudal traction), edema, vascularity, and surface tension properties. A review of these factors is beyond the scope of this article (see Refs. 21, 46, 56, 60, 85, 88, 94, 97, 108, 109, 113, 127, 135).

The ability of the pharynx to conduct air is dictated by the behavior of collapsible tubes. When negative pressure is applied to one end of a collapsible tube, such as during spontaneous breathing, flow increases as a function of the applied pressure up to a point (P_{MAX}). Further increases in the magnitude of negative pressure produce no further increases in flow [maximum flow; V_{MAX} (43, 103)] and, under some circumstances, result in a paradoxical reduction in flow [negative effort dependence (43)]. Negative effort dependence is frequently seen during obstructive hypopneas as a progressive reduction in flow during the inspiratory phase of individual breaths, even though the driving pressure for airflow is increasing (18, 41). That flow becomes independent of, or negatively affected by, increasing effort clearly distinguishes the pharyngeal resistor from simple inspiratory resistors in which flow increases progressively as a function of applied pressure. This behavior also means that increased activity of pump muscles (e.g., diaphragm) cannot increase flow in this disorder. Thus, once an obstructive event develops, an increase in flow can occur only through changes in pharyngeal dilators’ activity.1

Another important consequence of this behavior is that the ventilatory response to the changes in P_{CO2} and P_{O2} that occur during the event is no longer determined by the response of pump muscles but by the response of pharyngeal dilators to these and other stimuli, and by the ability of these dilators to stiffen/dilate the pharynx.

Maximum flow in collapsible tubes is determined by tube area (A), tube compliance (ΔA/ΔP), and gas density (ρ) according to the following equation [tube law (20, 123)]:

$$V_{MAX} = A \times \left[ \frac{A}{\rho (\Delta A/\Delta P)} \right]^{0.5}$$

Thus V_{MAX} is directly related to tube area and inversely related to tube compliance. Because as luminal pressure increases, pharyngeal area increases and its compliance simultaneously decreases (Fig. 1A), the relation between luminal pressure and V_{MAX} is not so nonlinear and is, in fact, close to a linear relation over the flow range relevant to sleeping subjects (12, 86, 98) (see footnote 2 for a discussion of tube law vs. the

---

1 In theory, V_{MAX} can increase through mechanisms that do not involve recruitment of pharyngeal dilators, such as increased caudal traction associated with increases in lung volume (108), or reduction of inspiratory effort when negative effort dependence exists (43). However, once an obstructive event is initiated, neither of these other mechanisms is available to increase V_{MAX} since lung volume cannot increase and inspiratory effort will increase rather than decrease. Thus the only mechanism available to increase V_{MAX} is recruitment of pharyngeal dilators.
Starling resistor model). This greatly simplifies the situation since the mechanical properties of an individual’s pharynx can be described simply by an intercept (P_Close) and a slope of the pressure-V_{\text{MAX}} relation (12, 86, 98). Early determinations of this relation were made by measuring V_{\text{MAX}} at different levels of constant sustained airway pressure (29, 100, 103). Because compensatory mechanisms were allowed to evolve at each pressure level before V_{\text{MAX}} was measured, the reported slope and intercept did not characterize the passive pharynx but, rather, the combined passive and active properties of the pharynx. Such measurements, therefore, could not be used to identify the separate roles played by passive (i.e., structural) properties and control mechanisms in the pathogenesis of OSA. More recently, measurement of V_{\text{MAX}} during the first few breaths after a reduction of continuous positive airway pressure (CPAP) (dial down) has been used to advantage to determine the passive mechanical properties of the pharynx (12, 86, 98, 127–129). This approach is based on the observations that on CPAP, dilator activity is minimal (57, 88, 95) and remains unchanged for 2–3 breaths following reduction of CPAP (62, 91, 98, 106). Thus, by briefly reducing CPAP to different levels and observing V_{\text{MAX}}, one can conveniently estimate the relation between external P_{AW} and V_{\text{MAX}} in the passive pharynx with reasonable accuracy (91, 98, 127).

Figure 1B shows a range of P_{AW}-V_{\text{MAX}} relations observed in sleeping subjects in whom the brief dial-down approach was used (e.g., 98, 127). Consistent with endoscopic measurements in the passive pharynx (e.g., 46), P_{Close} ranges from less than –10 to 10 cmH_{2}O (127). The slope, which reflects the shape of the pressure-area relation (see footnote 2), is also highly variable, ranging from 16 to 103 l s^{-1}cmH_{2}O^{-1} (127). A particularly useful value obtained from such plots is V_{\text{MAX}} at an external airway pressure of zero (V_{\text{MAX}0}, Fig. 1B). V_{\text{MAX}0} indicates what maximum flow would be during unassisted breathing if pharyngeal dilator influence declined to zero during sleep. Accordingly, this term incorporates the combined effects of P_{Close} and the slope of the relation in the nearly passive pharynx and, further, takes into account any pressure drop between the external airway and the site of flow limitation (i.e., upstream or nasal resistance). In essence, it embodies in one value the effect of all the anatomic factors that determine susceptibility to OSA (127).

An additional advantage of V_{\text{MAX}0} is that it can be very rapidly determined from a few brief dial down(s) to near atmospheric pressure. If the dial downs to atmospheric pressure result in complete obstruction, a few more dial downs to higher pressures may be added to determine the value of P_{Close} (e.g., 129). This simple approach is to be contrasted with the multiple observations over a wide range of pressures (including negative pressures) required previously to determine P_{Close} and the slope (12, 29, 86, 89, 98, 100, 103). V_{\text{MAX}0}, so determined, is also fairly consistent within a given patient/body position combination (127). By being so easily and rapidly obtained, it allows a bottom-line assessment of passive pharyngeal properties that can be readily and repeatedly performed in standard clinical laboratories (e.g., 127). This should make it possible to evaluate passive mechanics under different conditions in the same night (e.g., different body positions, sleep stages, time of night, and the effect of various pharmacological or mechanical interventions).

RELATION BETWEEN PASSIVE PHARYNGEAL PROPERTIES AND POLYSOMNOGRAPHIC SEVERITY OF OSA

There are only a few studies that evaluated this relation (22, 45, 60, 127). In other studies (e.g., 29, 32, 100) pharyngeal mechanics were not determined under passive conditions. In all four relevant studies (22, 45, 60, 127) there was a significant positive relation between apnea-hypopnea index (AHI) and some index of collapsibility of the pharynx. However, the relation showed considerable scatter in all cases. Pressure-area data obtained under anesthesia/paralysis by Isono et al. (45) showed considerable overlap among patients with severe, mild, or no OSA. When P_{Close} (45) or P_{\text{CRIT}} (60) was used as the index of collapsibility, it accounted for only 14% to 21% of the variability in the respiratory disturbance index (r^2 = 0.14 and 0.21 in the 2 studies, respectively). Although with use of the more comprehensive descriptor of passive properties (V_{\text{MAX}0}) the relation was stronger, much of the variance in the AHI was still unrelated to passive pharyngeal mechanics (r^2 = 0.28 for V_{\text{MAX}0} vs. AHI; Ref. 127). These findings indicate that the response to the abnormally collapsible pharynx plays a more important role in determining the polysomnographic picture than the severity of the anatomic problem. In the following sections the response to the obstructive events, and how this may affect the polysomnographic features, will be discussed.

COMPENSATORY MECHANISMS AVAILABLE TO RESTORE VENTILATION FOLLOWING A SLEEP-INDUCED OBSTRUCTIVE EVENT

Even the mildest hypopnea, conventionally defined as a 50% reduction in breathing amplitude (1) [which the author consid-
ers as 50% reduction in tidal volume (VT), would result in near doubling of arterial PCO₂ (PaCO₂) and severe desaturation unless some compensatory mechanism is engaged to restore ventilation. More significant events would be inconsistent with life in the absence of such compensatory mechanisms. One very important compensatory mechanism is arousal from sleep (90, 93). Arousal removes the precipitating cause of the obstruction and is associated with a substantial increase in dilator activity (10, 14, 61, 93), thereby restoring ventilation. Although arousal is an effective safeguard against life-threatening asphyxia, such a compensatory mechanism has the disadvantage that the obstruction will recur once the subject returns to sleep. Thus, without arousal-independent mechanisms, even the patient with the mildest form of anatomic compromise would be doomed to a constantly interrupted sleep.

In order for arousal-independent mechanisms to succeed, they must be able to restore minute ventilation (Ve) to a sustainable level without arousal. A sustainable Ve is one with which the changes in blood gas tensions in the steady state are not enough to trigger arousal. Sustainable Ve is between 80 and 95% of the unloaded ventilatory demand (see online supplement in Ref. 127) with the range largely reflecting how much change in gas tensions can be sustained without arousal (arousal threshold). Unloaded ventilatory demand during sleep (e.g., Ve on optimal CPAP) varies among individuals [3.8–9.3 l/min (127, 129)], reflecting differences in metabolic rate, dead space ratio, and CO₂ set point. Arousal threshold also varies greatly among patients (5, 6, 54, 80, 99, 112, 121, 129, 136) and at different times in the same patient (6). Thus, for a given abnormality in passive mechanics (e.g., VMAX0), the challenge facing the compensatory mechanisms varies between and within subjects.

Ve is the product of mean inspiratory flow rate (VT/Ti) and inspiratory duty cycle (Ti/Ttot) (73, 96):

\[ Ve = 60[(VT/Ti) \times (Ti/Ttot)] \]  

(2)

It follows that when an obstructive event develops, Ve may be increased toward the sustainable level through either an increase in duty cycle or an increase in mean inspiratory flow rate. In the presence of pharyngeal flow limitation, the latter can be accomplished only through activation of pharyngeal dilators. These two avenues will now be discussed.

Compensation via an Increase in Inspiratory Duty Cycle

With sustained flow limitation, Ti/Ttot increases (47, 96, 104, 112, 121). An increase in duty cycle has the effect of increasing ventilation in the absence of any increase in VMAX since the patient spends a greater fraction of time inspiriting (Eq. 2). Figure 2 is an example of how effective this can be. While the patient was on CPAP (Fig. 2, left), peak flow was 0.48 L/s and Ve was 7.2 l/min. On dial down of CPAP, flow became limited and VMAX was 0.18 l/s, or 37% of peak unloaded flow. VT decreased from 0.46 to 0.20 liter, or 43% of eupneic VT. Because respiratory rate did not change, the immediate decrease in Ve was also to 43% of the eupneic value. Ti/Ttot increased gradually from 0.38 immediately after dial down (similar to the value on CPAP) to 0.61 a few minutes later (Fig. 2, right) while VMAX remained the same (37% of peak unloaded flow). As a result, Ve increased from an initial 3.1 l/min at the onset of hypopnea to 6.0 l/min later on, reaching 85% of baseline Ve on CPAP. This was associated with only a 5-mmHg increase in end-tidal PCO₂ (PetCO₂) and a 2% reduction in O₂ saturation. These values are consistent with a steady state. The change in chemical drive was clearly tolerated by this patient as evidenced by continued sleep. Thus in this case a VMAX of only 37% of peak unloaded flow was sustainable without arousal.

The above example helps illustrate the important point that for a given peak unloaded flow, sustainable VMAX and, hence, how much flow limitation is consistent with stable sleep are very much a function of how much Ti/Ttot can increase with sustained loading. The increase in Ti/Ttot during sustained loading is the result of two mechanisms. First, neural inspiratory time (Ti) (e.g., measured from diaphragm EMG) increases and neural expiratory time (Te) decreases as the load is sustained (81, 104, 112, 121). These neural changes are linked to the increase in inspiratory effort (81, 104, 112, 121) and may be mediated by upper airway mechanoreceptors responding to the greater deformation associated with the more negative pharyngeal pressure (81). The response of neural Ti/Ttot to increasing effort varies considerably among patients (96, 104), and there is evidence in mice that the duty cycle response to hypercapnia is, in part, genetically determined (96). Second, with increased effort, the duration of maximum inspiratory flow would increase even if neural Ti did not. This is because when inspiratory effort increases, the fraction of neural Ti during which intrathoracic pressure is more negative than PMAX is higher (Fig. 2, inset). The impact of this mechanism depends on the shapes of the rising and declining phases of the effort, and both of these vary considerably among subjects (133). Furthermore, since both neural and mechanical mechanisms of increasing Ti/Ttot are related to the increase in effort, the increase in Ti/Ttot is ultimately dependent on how much increase in effort can be reached before arousal occurs [i.e., arousal threshold (3, 30, 54)]. This, also, varies considerably among subjects (5, 6, 54, 80, 99, 112, 121, 129, 136). This wide range of Ti/Ttot response translates into a wide range of sustainable VMAX when the latter is expressed as a percentage of peak unloaded flow. In this author’s experience, sustainable VMAX ranges from >70% peak flow (i.e., recurrent arousals occurring at VMAX of 70% peak) to as low as 30% peak unloaded flow (e.g., VMAX during steady snoring in some patients), with the example shown in Fig. 2 representing nearly the most that prolongation of duty cycle can accomplish in this respect.

Because peak unloaded flow varies greatly among subjects [0.17–1.08 l/s (127)] and the fraction of peak flow that can be sustained also varies widely, sustainable VMAX varies greatly when expressed in actual flow units. For example, in a patient with a high ventilatory demand (e.g., peak unloaded flow = 0.7 l/s) and a low arousal threshold (sustainable VMAX = 70% peak unloaded flow), sustainable VMAX may be as high as 0.5 l/s. By contrast, for a patient with low ventilatory demand (peak unloaded flow = 0.20 l/s), a high arousal threshold, and favorable inspiratory effort shape (sustainable VMAX = 40% peak unloaded flow, e.g., Fig. 2), sustainable VMAX may be < 0.1 l/s. Furthermore, since some of the factors that determine sustainable VMAX change during the night in the same individual [e.g., arousal threshold (2)], sustainable VMAX may also vary from time to time during sleep.
Peak unloaded flow and sustainable $\dot{V}_{\text{MAX}}$ of the patient illustrated in Fig. 2 are inserted in Fig. 1B to illustrate how these two variables interact with passive mechanics to determine the polysomnographic outcome in the absence of compensatory increase in $\dot{V}_{\text{MAX}}$. For a patient with these “control” characteristics, the anatomy represented by line A in Fig. 1B would be inconsequential. The anatomy represented by line B in Fig. 1B would result in moderately severe flow limitation, but a steady state can be reached without any active increase in $\dot{V}_{\text{MAX}}$. In a patient with the same “control” characteristics, an anatomy represented by lines C or D of Fig. 1B would be inconsistent with stable breathing unless $\dot{V}_{\text{MAX}}$ increased to match sustainable $\dot{V}_{\text{MAX}}$ through appropriate activation of pharyngeal dilators [see effective recruitment threshold (T$_{ER}$) below]. Clearly, the impact of the same anatomic abnormality would be different in a patient with different control characteristics. For example, line C or D of Fig. 1B may be associated with stable flow limitation in a patient whose sustainable $\dot{V}_{\text{MAX}}$ is 0.1 l/s and, conversely, line B of Fig. 1B may be incompatible with steady breathing in a patient whose sustainable $\dot{V}_{\text{MAX}}$ is $>0.3$ l/s. In a recent study (127), ventilatory demand on CPAP was found to be an independent determinant of the AHI. The higher AHI in obese subjects is in part related to their higher flow demand (127).

When $\dot{V}_{\text{MAX}}$ is greater than sustainable $\dot{V}_{\text{MAX}}$, steady breathing would continue so long as arousal threshold remains high. A decrease in arousal threshold would increase sustainable $\dot{V}_{\text{MAX}}$ and may make it impossible to reach the level of chemical drive required for a steady state (i.e., that associated with a high enough Ti/Ttot). Arousal would then occur. The associated increase in ventilation would reduce chemical drive, returning Ti/Ttot and Ve to a lower level, and the cycle repeats. Thus either steady breathing or recurrent hypopneas would follow depending on arousal threshold.

It can also be appreciated that if the respiratory neural responses to CO$_2$ and hypoxia are low, the changes in PaCO$_2$ and arterial Po$_2$ (PaO$_2$) required to increase inspiratory effort sufficiently for a steady state to occur through the duty cycle mechanism may be quite large. Because arousal is related to the intensity of effort and not to blood gas tensions, per se (30, 54), the patient may sleep through substantial deterioration of blood gas tensions, particularly if arousal threshold is also high. Such a combination (collapsible pharynx, low chemosen-
sitivity, high arousal threshold) may underlie the obesity-hyperventilation syndrome.

Whereas the increase in duty cycle makes it possible to sustain a lower $V_{\text{MAX}}$ than would otherwise be the case, it is clear that this mechanism can, by itself, be effective only in dealing with relatively mild anatomic problems. Where $V_{\text{MAX}}$ is less than sustainable $V_{\text{MAX}}$, including all cases where $P_{\text{Close}}$ is positive, stable breathing can only occur through active dilatation of the pharynx. This introduces another level of complexity in that breathing stability is now dependent on how effective the control system is in raising $V_{\text{MAX}}$ and in keeping it high in a stable fashion.

Compensation via Reflex Increase in $V_{\text{MAX}}$

Possible mechanisms for increasing $V_{\text{MAX}}$. As indicated earlier, once an obstructive event is initiated the only way to increase flow is through increased activation of pharyngeal dilators. There are only two known mechanisms that can increase dilator activity during sustained sleep.

I) Negative Pharyngeal Pressure. Negative pressure in the upper airway increases activity of the genioglossus in anesthetized animals (68) and in awake humans (38, 72, 118, 119). The gain of the reflex is reduced during sleep (37, 72, 118, 119). Nonetheless, there is evidence that this reflex contributes to activation of the dilators as chemical drive increases in the course of sleep-induced obstructive events (10, 77). Whether this reflex is weaker in OSA patients is unclear since available results are contradictory and limited to the awake state (11, 75).

II) Increase in Chemical Drive. In sleeping animals, genioglossus activity increases in response to CO$_2$ inhalation (35, 39, 131). However, a finite increase in chemical drive above eupneic level must occur (threshold) before genioglossus activity increases (35, 39, 131). In awake normal subjects (82, 83), and awake OSA patients (77), genioglossus activity increases in response to CO$_2$ and hypoxia without an apparent threshold. Studies on the response of dilators to increased chemical drive during sleep in normal subjects have yielded conflicting results. In two studies (92, 105), hypoxia and hypercapnia did not increase genioglossus activity over the range of hypercapnia and hypoxia that is possible before arousal (50–70% increase in $V_{\text{E}}$), thereby suggesting the presence of a threshold. However, other studies do not show such a refractory range (65). The reason for these discrepancies among normal subjects is unclear.

In most OSA patients severe flow limitation or complete obstruction develops on a dial down from CPAP (127). Yet, dilator activity does not increase immediately (62, 91, 98, 106). This indicates that the negative pressure associated with eupneic drive (i.e., on CPAP) is insufficient to augment dilator activity. Pharyngeal pressure can become more negative only if pump muscles work harder, and, during sleep, this can only occur if chemical drive increases (126). Accordingly, an increase in chemical drive is necessary to increase dilator activity regardless of whether the immediate stimulus is chemical drive itself or the associated increase in negative pressure. This is a fundamental observation in that it indicates that for $V_{\text{MAX}}$ to increase without arousal, chemical drive must increase above the eupneic level and remain elevated above the level required to generate a sustainable $V_{\text{MAX}}$. If drive increases above this level but decreases again below it when the airway opens, recurrent obstructions would result.

Will an increase in dilator activity without arousal necessarily increase $V_{\text{MAX}}$? Dilator activity increases progressively in the course of obstructive events (10, 15, 61, 77, 84, 93). Yet, $V_{\text{MAX}}$ does not increase, and often decreases, for some time despite the clear increase in this activity. Remmers et al. (93) proposed their enduring “balance of forces” theory to explain this paradox. According to this mechanism, the increase in dilator action as chemical drive increases is more than offset by the concomitant increase in negative pharyngeal pressure tending to collapse the airway. Thus the balance of forces during chemical activation of the dilators is in favor of keeping the airway closed and remains so until a preferential activation of the dilators occurs through arousal. However, a number of recent observations indicate that activation of the dilators via an increase in chemical drive can prevail over the collapsing negative pressure if chemical drive is allowed to increase sufficiently. These observations will now be summarized:

i) Most patients with OSA develop periods of stable breathing and sleep (127). By itself, such a finding might be attributed to mild anatomic problems that can be offset part of the time without an increase in $V_{\text{MAX}}$ (for example as discussed in relation to Fig. 2). However, Younes (127) found that 78% of patients with OSA develop periods of stable breathing and sleep even though their $V_{\text{MAX}0}$ was well below sustainable $V_{\text{MAX}}$ and was frequently zero (complete obstruction). Furthermore, the values of $V_{\text{MAX}0}$ obtained over several hours were sufficiently stable to rule out the possibility that alternation between stable breathing and OSA was related to within-night changes in passive mechanics. Thus most patients currently diagnosed with OSA are capable of activating their dilators enough to increase $V_{\text{MAX}}$ to sustainable levels without the benefit of arousal.

ii) The temporal relation between arousal and upper airway (UA) opening was examined in detail in the same patients (128). There was no arousal at or beyond opening in 17% of obstructive events (type 1 response) while in another 21% of events arousal followed, rather than preceded, UA opening (type 2 response). In either case, lack of arousal at the time of opening was ruled out by visual inspection and spectral analysis of the EEG, and by lack of heart rate increase up until UA opening.3 In the remaining 62%, arousal occurred at or before UA opening (type 3 response). The response type was not fixed in individual patients. Rather, most patients demonstrated two, or all three, response types at different times. The frequency of type 1 response (no arousal) increased, and that of type 3 decreased, as sleep depth increased (as indicated by delta power). Thus the relation between arousal and UA opening appeared to be incidental rather than a cause and effect relation. Even more importantly, the latency to UA opening (i.e., event duration) was the same whether arousal occurred before or after UA opening or did not occur at all (128).

iii) Where UA opening occurred without or before arousal, the increase in flow was greater than that required to restore stable

---

3 An increase in heart rate may occur beyond UA opening in the absence of cortical arousal, but this is to be expected as a physiological response to the large hemodynamic and lung volume changes following resolution of the obstruction (comparable, for example, to sinus arrhythmia).
ventilation (128). Thus arousal is not needed to increase $V_{\text{MAX}}$ adequately.

From these observations, it was concluded that in most patients arousal is not necessary for adequate compensation but that arousals occurred because the increase in chemical drive required for reflex opening (henceforth called $T_{ER}$) overlaps with the range of increase in chemical drive that triggers arousal in these patients.

How much must chemical drive increase to reflexly open the airway ($T_{ER}$)? Figure 3 shows examples of responses to dial down from air breathing in three OSA patients in whom arousal threshold was high. Here, chemical drive at the onset of the event was, by definition, eupneic drive. In all cases the initial response was complete obstruction. In patient A, the airway opened spontaneously without arousal on breath 3. Since the lung-to-carotid circulation delay is $\approx 2$ breaths (129), it can be concluded that in this patient the airway can open reflexly as soon as chemoreceptors sense any deterioration in blood gas tensions relative to eupnea. Thus $T_{ER}$ is essentially zero. In patient B, four obstructed breaths elapsed with no increase in flow despite a substantial increase in respiratory drive, as judged by the Respitrace signal. Flow appeared, with flow limitation, in breath 5 without arousal and increased much more in the next breath, still without arousal. In this case, $T_{ER}$ was clearly $>0$ but was still less than arousal threshold. In patient C, flow did not appear until arousal occurred despite five breaths of increasing efforts. During another dial down when sleep was deeper, the airway opened, still at breath 6, but without arousal (not shown). Clearly, if in these patients arousal threshold were lower (e.g., arousal occurring during breaths 3 or 4), as it usually is (see below), opening by arousal

4 In the presence of complete obstruction (i.e., no bulk flow into the chest), Respitrace amplitude reflects gas decompression and chest deformation related to the magnitude of negative intrathoracic pressure.

Fig. 3. Response to a dial down from optimal CPAP in 3 patients with a relatively high arousal threshold. In each case, the dial-down pressure was set to be just below critical closing pressure ($P_{\text{com}}$), thereby resulting in a complete obstruction. C3/A2, central electroencephalogram. $P_{AW}$, airway pressure. Patient A: opened the airway (arrow) without arousal after only 2 obstructed breaths. In patient B, the airway remained closed for 4 breaths despite increasing inspiratory efforts, as judged by the Respitrace signal. The airway opened with moderate flow limitation at breath 5, and flow increased much more at breath 6. Patient was in delta sleep throughout. While in stage 2 sleep, patient C remained obstructed for 5 breaths despite the marked increase in inspiratory effort. The airway opened following an arousal (heavy arrow). Note the different time scales in the 3 panels.
would have preempted reflex opening in patients B and C, making it appear that arousal was the responsible mechanism.

From these examples (Fig. 3) it is clear that \( T_{ER} \) varies considerably among patients. In a recent study (129) the actual value of \( T_{ER} \) was measured in 21 patients with severe OSA (AHI = 91 \( \pm \) 24 h\(^{-1}\)) by increasing chemical drive (i.e., \( V_E \) on CPAP) to different levels before dial downs. \( V_{MAX} \) during the dial down was then plotted against pre-dial-down \( V_E \) (Fig. 4). The minimum increase in pre-dial-down \( V_E \) associated with an effect an increase in \( V_{MAX} \) (i.e., \( T_{ER} \)). Figure 4 shows the spectrum of the results obtained. The absissa value of the rightmost point in each panel represents the highest \( V_E \) that could be reached without arousal. In 12 patients (57%) the arousal threshold permitted a determination of \( T_{ER} \). Figure 4, A–C, represents the highest \( V_E \) observed in this group. The increase in \( V_E \) required for effective recruitment ranged from 0 l/min (Fig. 4A; patient A in Fig. 3) to 9 l/min (109% of eupneic \( V_E \); Fig. 4C), with an average of 74 \( \pm \) 28% baseline. In the remaining nine patients, \( T_{ER} \) was greater than what was permitted by the arousal threshold (e.g., Fig. 4D; \( T_{ER} > 174\% \) baseline; patient C in Fig. 3). In several of these 9 patients the difficulty was not necessarily that \( T_{ER} \) was higher than in the other 12 patients, but that the range of ventilatory stimulation that could be attained without arousal was very narrow [permissible range of increase in \( V_E \) varied from 52 to 174% baseline (129)].

It follows that the majority of patients are able to maintain the airway open without arousal with modest increases in chemical drive, corresponding to an increase in unobstructed \( V_E \) of \(<9.0 \) l/min (5.0 l/min on average). In the same study (129) it was determined that such increases in unobstructed \( V_E \) (i.e., on CPAP) could be accomplished, on average, within a few breaths of increasing \( PETCO_2 \) 3.3 mmHg along with a 4% reduction in \( O_2 \) saturation. Given that even smaller changes would be required to increase chemical drive by the same amount in the steady state (19), the steady-state changes in gas tensions required to mount effective compensation in most patients are quite modest.

It may be argued that \( T_{ER} \) determined by the above method underestimates the increase in drive required to open the airway from an already obstructed state. This may be so because the pressure required for opening an already closed airway is higher than the pressure at which the airway closes starting from an open position (59, 79, 109, 122). However, if such a difference exists it cannot be large. Thus, in an earlier study (128), 59 of 82 patients (72%) opened the airway without arousal following at least one dial down (128). These dial downs started at eupneic respiratory drive (air breathing on CPAP) (128). In half these cases opening was from a completely obstructed state while in the other half the dial down was associated with hypopnea. The latency to airway opening was only 16.5 \( \pm \) 8.0 s (range 6–43 s), and there was no difference between events with hypopnea and complete obstruction (16.5 \( \pm \) 8.2 vs. 16.5 \( \pm \) 7.9). Considering that 1) these dial downs began from eupneic drive, 2) all events were terminated before recirculation, and 3) the decrease in \( O_2 \) saturation was on average only 3% (128), the increase in drive at the time of opening (on average 7 s beyond the circulatory delay) could not have been much different from \( T_{ER} \) determined by increasing chemical drive before dial down of CPAP (e.g., Fig. 4).

**Why is it necessary for chemical drive to increase a threshold amount (\( T_{ER} \)) before opening the airway, and why is \( T_{ER} \) so different among patients?** When the airway is completely obstructed (i.e., \( P_{Close} > 0 \)), particularly if the tongue is impacted in the oropharynx, it is easy to see how, as postulated by Remmers et al. (93), the increase in suction pressure as chemical drive rises counteracts the increase in dilating force produced by dilators’ activation. Thus, opening can only occur if the dilators generated enough force to counteract the rising pump muscle pressure (which results in more negative pharyn-

---

**Fig. 4.** Effect of increasing chemical drive before dial down on flow during dial down in 4 representative patients (A–D). ○, Dial-down flow when patient breathed air on CPAP. Note severe flow limitation in all cases. Dotted horizontal line, mean \( \pm \) 2SD of the response during air dial downs. ●, Dial-down flow when chemical drive was increased to different levels before dial down (absissa) by changing inspired gas concentrations. \( T_{ER} \), effective recruitment threshold, the increase in chemical drive above which dial-down flow became sensitive to further increases in drive. Note that different levels of stimulation were required in patients A–C to effectively dilate the airway, while in patient D maximum flow during dial down was unchanged despite a 3-fold increase in chemical drive. Modified from Ref. 129.
geal pressure), plus any additional opening pressure requirements dictated by the passive mechanical properties, namely, $P_{\text{Close}}$ and the viscous forces that resist opening (57, 58, 109, 122). This would clearly require the force generated by dilators to increase at a faster rate than the rate of increase in the collapsing negative pharyngeal pressure, thereby ultimately prevailing over it. It was initially felt that such a relatively greater activation of the dilators could only occur with arousals. The recent finding that in most patients the dilators can prevail over the opposing forces without arousal if chemical drive increases sufficiently (127–129) indicates that, in most patients, dilator force increases at a faster rate than the rate of increase in pump muscle pressure. Thus, when $P_{\text{Close}}$ is $>0$, $T_{\text{ER}}$ is likely a function of $P_{\text{Close}}$, viscous properties of the pharynx, and the relative rates of recruitment of pump and dilator muscles. To the extent that all these variables can differ among patients, $T_{\text{ER}}$ can be quite different.

The situation is somewhat harder to explain when $P_{\text{Close}}$ is $<0$. Here, the most that can happen is a hypopnea with flow limitation. In theory, with flow limitation, downstream pressure (i.e., the negative pressure generated by pump muscles) should be irrelevant and any dilating force applied at the site of flow limitation should increase $V_{\text{MAX}}$. Several possibilities may explain failure of flow to increase during hypopneas despite the progressive increase in respiratory drive.

i) As in sleeping animals (35, 39, 131), a threshold increase in chemical drive may be needed before dilator activity can increase. Data in sleeping normal subjects are conflicting (65, 92, 105), and whereas many studies have documented a progressive increase in dilator activity during obstructive events in OSA patients (10, 15, 77, 83, 93), the chemical drive, relative to eupnea, at which recruitment begins in these patients has not been determined.

ii) There are many dilator muscles that differ in their site of action and their pattern of activity and response to different stimuli (36, 48, 111, 120). The site of flow limitation differs among patients (74). Most data on dilator muscles’ activation were derived from easily accessible muscles (e.g., genioglossus). The specific dilator muscle(s) relevant to the site of flow limitation in a given individual may have different response characteristics from those that have been recorded from (48) and may require a higher chemical threshold for activation.

iii) Even though the pharynx is not completely closed, expanding it may require a threshold dilating force to overcome “stickiness” of apposed mucosal folds.

iv) As mentioned earlier, negative effort dependence may occur during flow limitation (43). In such cases, the ability of the dilators to increase $V_{\text{MAX}}$ as chemical drive increases is compromised by the increasing negative pressure, tending to reduce $V_{\text{MAX}}$. Thus, when negative effort dependence is present, $V_{\text{MAX}}$ can increase only when the increase in dilator force can cancel out the negative effect of increasing effort on $V_{\text{MAX}}$. This may not occur until a threshold increase in chemical drive. The presence and magnitude of negative effort dependence appear to vary among OSA patients (43).

v) Regardless of the immediate reason for a $T_{\text{ER}}$ (see i–iv, above), and whether the event is a hypopnea or apnea, an increase in $V_{\text{MAX}}$ can only occur when the force generated by the dilators prevails over opposing mechanical forces. The increase in dilator force as chemical drive increases is a function of the relation between chemical drive and dilator activity. This relation is clearly affected by the gain of the negative pressure reflex. With a stronger reflex, dilator activity should rise faster as chemical drive, and pharyngeal negative pressure, increase, and this should result in dilator force prevailing at a lower chemical drive (i.e., lower $T_{\text{ER}}$). In this respect it is noteworthy that the function of some pharyngeal mechanoreceptors that likely mediate the negative pressure reflex (36) is impaired in OSA patients (33, 55, 76). Thus differences in the potency of the negative pressure reflex may account, in part, for differences in $T_{\text{ER}}$ among patients.

vi) The rate of increase in dilator force as chemical drive increases is also affected by the relation between dilator activity and the force generated by the muscles (i.e., dilators’ contractile properties). Thus, if the dilators were stronger, a lesser degree of activation (i.e., lesser chemical drive) would be required to prevail over the opposing mechanical forces.

Inflammation and denervation (13) as well as impairment in the contractile properties (16) of pharyngeal dilators have been reported in OSA patients. Accordingly, differences in dilator muscles’ contractile properties may contribute to differences in $T_{\text{ER}}$ among patients.

Recently, Patil et al. (87) compared OSA patients and normal subjects with respect to their ability to decrease critical closing pressure ($P_{\text{CRIT}}$) during sustained reductions in airway pressure. OSA patients failed to decrease $P_{\text{CRIT}}$ whereas normal subjects did. This suggests that $T_{\text{ER}}$ is higher in OSA patients than in normal subjects. Interestingly, the response of the genioglossus was not different, suggesting that the higher $T_{\text{ER}}$ in OSA patients may be related to a mechanical threshold (see iii and iv, above) or to impaired conversion of electrical activity to mechanical force (see mechanism vi, above). Further studies are clearly needed to elucidate the mechanism of $T_{\text{ER}}$.

MECHANISMS OF INSTABILITY IN OSA

If modest increases in chemical drive can open the airway without arousal, why do patients with OSA develop recurrent obstructions? For obstructive events to be prevented it is necessary that chemical drive remain continuously above the level required to obtain sustainable $V_{\text{MAX}}$. For this to happen, flow cannot increase much above sustainable $V_{\text{MAX}}$ following airway opening. The occurrence of a significant ventilatory overshoot would wash away the chemical stimulus required to keep the airway open, setting the stage for another obstruction. Yet, OSA patients routinely develop an overshoot at the end of obstructive events. A recent study found that the increase in flow at airway opening was $2.56 \pm 1.69$ times the amount required to reach sustainable $V_{\text{MAX}}$ (128). Thus the reason why OSA patients develop recurrent obstructions when they do is that their ventilation overshoots the target on airway opening. Accordingly, as has been suggested earlier (67), OSA is an example of control system instability. The general mechanisms of ventilatory instability have been a subject of much study and modeling and have been reviewed extensively (17, 50, 53, 67, 125, 130). Much of the experimentation and modeling concerned central apneas (Cheyne-Stokes breathing). However, as has been theoretically suspected (e.g., 52, 67, 125), the instability of OSA differs from the usual periodic breathing in several important respects. i) The ventilatory response to the blood gas changes that occur during and following obstructive event (i.e., controller gain) is not governed by the response of
pump muscles and chest mechanics, which determine the classic ventilatory responses. Rather, it is governed primarily by the response of upper airway dilators and upper airway mechanics. Unlike the usual ventilatory responses, these are discontinuous responses (see Fig. 4), and the slope may be much steeper than the ventilatory responses that would obtain in the same individual with an open airway (125). 2) Arousals typically occur at the end of the event, thereby introducing additional complexity (50, 52, 67, 125). 3) Cycling frequency in not simply a function of circulatory delay, as is the case with simple periodic breathing, but is primarily related to the time it takes for the upper airway to open (see below). In the following account recent experimental data that have specifically addressed the mechanisms of the ventilatory overshoot (i.e., instability) in OSA will be reviewed.

Occurrence of Unnecessary Arousals

The occurrence of arousals in patients in whom the obstruction continues despite dangerous changes in blood gas tensions (i.e., very high T ER) is unquestionably necessary and beneficial. However, and despite the fact that arousal threshold is higher in OSA patients than in normal subjects (3), in most patients who are currently diagnosed with OSA, arousal occurs in response to minimal, nonthreatening changes in gas tensions. In 82 OSA patients with mixed severity (AHI = 46 ± 35 h⁻¹) the average time to arousal following a CPAP dial down (i.e., beginning with eupneic blood gas tensions) was only 17.7 ± 11.0 s (128). At this time, O₂ saturation had decreased by only 3.1 ± 3.5% and PCO₂ at the chemoreceptors could not have increased by more than a few millimeters Hg above eupnea (128). In a more recent study (129), arousal threshold, defined as the increase in Ve above eryptna associated with arousal 50% of the time, was <10 l/min (6.3 ± 2.3 l/min) in 15 of 21 patients with very severe OSA. In the same patients, such increases in Ve could be produced within a few breaths of increasing the PCO₂ by 3.3 mmHg along with 4% decrease in O₂ saturation (129). Furthermore, because in most patients T ER is reached with only modest blood gas changes, airflow opening would still occur at very safe blood gas tensions in the absence of arousals (128). Thus, in a majority of patients, arousals are not needed and occur simply because arousal threshold is unnecessarily low.

Arousals dramatically increase the ventilatory overshoot in OSA patients (128). The effect of arousals on postevent flow is not all-or-none but varies with the intensity of arousal (Fig. 5). With barely noticeable arousals, the overshoot is slightly higher than in the absence of arousal, whereas with the most intense arousals, postevent flow is nearly four times what is required for stable breathing (Fig. 5). Arousals also vary in duration. Since the reduction in chemical drive during the ventilatory phase depends on the duration as well as the amplitude of overshoot, it may be expected that arousals of longer duration would increase the likelihood of instability.

Arousal threshold varies greatly among OSA patients (5, 6, 54, 79, 99, 112, 121, 136). Furthermore, it also varies over a wide range across the night in the same patient (2). These differences likely explain in part why some patients develop recurrent OSA while others do not despite similar upper airway mechanics (127) and also, how the same patient may develop recurrent events at times during the night and have stable breathing at other times in the same body position (127).

In a recent study where both T ER and arousal threshold were determined in the same patients, and both were expressed as the percent increase in eupneic Ve required for either response to occur, the two thresholds were, as predicted earlier (128), in the same general range but there was no correlation between them (129), indicating that the two responses involve different central mechanisms.

Magnitude of T ER

The likelihood and extent of overshoot increase as T ER increases. Assume that T ER is 50% of eupneic drive. With an obstructive event the airflow would open reflexly when inspiratory effort (i.e., negative intrathoracic pressure) is only 50% greater than eupneic level. Under such conditions the maximum possible overshoot, which would occur if upper airway resistance decreased to normal levels during the ventilatory phase, is 50% above eupneic ventilation. If resistance is not completely normalized during the ventilatory phase, the overshoot would be even less. With such a low T ER, there would also be a smaller likelihood of arousal occurring during the ventilatory phase. Now, assume that T ER is 200%. Here the potential for a large overshoot, up to 200% of eupneic ventilation if resistance is normalized, is clearly much greater when the airflow opens and, since arousal is more likely to occur at such high drive levels, airflow opening is more likely to be complete and the potential for reaching the maximum possible overshoot is thus greater. In this author’s opinion, the magnitude of T ER is the single most important variable that impacts the tendency to oscillate in the presence of abnormal pharyngeal mechanics.
Flow Response in the First Breath After Airway Opening in the Absence of Arousal

Flow usually increases abruptly in the first breath following an obstructive event. This discontinuous flow behavior supported the notion that arousals are the mechanism for airway opening. Recently, the magnitude of increase in flow at airway opening was measured when arousal was not present at the time (128). Although less than when arousal occurs, the discontinuous increase in flow was still evident in the first breath in many cases (128). In 69% of cases, flow during the first breath was greater than flow on CPAP. There was a wide range of these first-breath flow responses. These findings indicate that in some patients there is a sudden, very substantial reduction in airway resistance once T_{ER} is reached, even in the absence of arousal. Such a development would clearly promote an overshoot.

The mechanism for this discontinuous increase in flow during arousal-free airway opening is not clear. It is likely related to the same mechanism responsible for T_{ER}. In particular, it could be related to a mechanical threshold, such as the need to overcome viscid secretions between apposed mucosal surfaces, or to prevail over negative effort dependence. In the former case, once the dilators manage to unstick the apposed surfaces, the airway would pop open with a sudden large increase in flow. In the latter case, once the dilators manage to generate a force that overcomes the negative effort dependence, flow will increase. As a result, pharyngeal pressure becomes less negative, which in the presence of negative effort dependence further increases flow, and so on. Flow would thus increase rapidly through positive mechanical feedback.

Flow and V̇e Beyond Breath 1

Typically, dilator activity decreases and upper airway resistance increases beyond the first open breath (15, 77). This is presumably because pharyngeal pressure is not as negative. A reduction in the intensity of arousal beyond breath 1 may also be a factor. This increase in resistance would tend to attenuate the overshoot. Mitigating these favorable changes is the tendency for chemical drive to continue increasing beyond the first open breath. This is because the blood gas tensions at the chemoreceptors at the time of opening reflect the gas tensions that existed in pulmonary blood some time before opening (i.e., lung-carotid circulation delay). In the interim, pulmonary gas tensions continue to deteriorate. Thus, depending on the transit time, chemoreceptors may continue to sense deterioration in gas tensions, and chemical drive may continue to increase, for a finite interval beyond breath 1. This would increase or prolong the overshoot because 1) arousal may be triggered after opening (e.g., later in breath 1 or in subsequent breaths) if it was not already present (128); 2) if arousal were present, its intensity and duration may increase; or 3) even without arousal the continued increase in chemical drive may attenuate or even reverse the increase in resistance that would otherwise occur beyond breath 1. The relation between chemical drive and V̇_{MAX} (and hence V̇e) beyond T_{ER} is highly variable and is often very steep (128; Fig. 4) so that even a small increase in drive beyond the point of opening may result in a large increase in flow and a much greater overshoot (e.g., Fig. 3B).

The magnitude of postevent increase in chemical drive depends on the lung-to-carotid circulation time and the rate of increase in chemical drive during this interval. The circulation delay was recently found to be 9.9 ± 2.4 s in patients with severe OSA (129). In the majority of patients (67%) this represented a two-breath delay, indicating that the increase in chemical drive would extend for one breath beyond breath 1. In 19% the delay was the duration of only one breath (i.e., no expected increase in drive beyond breath 1), while in the remaining 14% the delay would result in chemical drive increasing for two breaths beyond breath 1.

The rate of increase in chemical drive during and following an obstructive event is determined by 1) how much V̇e decreases during the event; 2) factors that determine the rate of change in pulmonary gas tensions per unit reduction in V̇e [so-called plant gain (17, 50, 53, 67, 125, 130), including lung volume, cardiac output, and metabolic rate; and 3) the dynamic response of the respiratory centers to these gas changes (i.e., controller gain; for reviews see 17, 50, 53, 67, 125, 130). There is much indirect evidence to suggest that, on average, for a given reduction in alveolar CO₂ the rate of increase in chemical drive is higher in OSA patients than in normal subjects. Thus:

1) Hudgel et al. (42) found the dynamic response of V̇e to a single breath increase in inspired CO₂ to be higher in OSA patients than in normal subjects. This study was, however, done in awake upright subjects breathing oxygen.

2) The use of proportional assist ventilation (PAV) superimposed on CPAP (to stabilize the airway) was recently introduced to measure loop gain during sleep (130). Loop gain is the product of plant gain (i.e., see 2 above) and controller gain (i.e., see 3 above) when respiratory mechanics are constant (50, 53). It is a dimensionless number that describes the ratio of plant gain (controller gain) over controller gain (plant gain) during sleep. When loop gain is >1, oscillations in breathing will not occur spontaneously. However, higher values in the range between zero and 1.0 reflect a greater rate of rise in chemical drive following a given reduction in V̇e, and a greater tendency to oscillate if other destabilizing factors, such as unstable mechanics or arousals, are added (130). In normal subjects loop gain is very low (usually <0.3; Refs. 69, 70, 116, 117). In patients with OSA, loop gain is almost always less than 1.0 (115, 130), indicating that, without additional destabilizing factors, OSA patients would not be unstable. However, loop gain measured on CPAP is, on average, higher in OSA patients than in normal subjects, indicating that, for a given reduction in V̇e, the rate of rise in chemical drive per unit reduction in V̇e should be higher in such patients. Wellman et al. (115) found only a weak correlation between loop gain and the AHI. This is not surprising since, as described above, loop gain measured on CPAP reflects only some of the factors that determine the rate of rise in chemical drive (another important factor being the severity of the hypopnea), and the rate of rise is itself only one of many factors that determine instability in OSA patients, the others being arousal threshold and intensity, magnitude of T_{ER}, relation between chemical drive and flow beyond T_{ER} (Fig. 4), and the discontinuous flow response at the time of opening. Interestingly, however, these authors (115) found a much stronger correlation between loop gain and AHI in a subgroup of patients in whom Ṗ_{Close} was near zero, thereby illustrating the interaction between severity of upper airway mechanics and chemical control instability.
3) The average rate of increase in negative intrathoracic pressure (ΔP/Δt) during naturally occurring obstructive events in OSA patients ranged between 1.0 and 2.0 cmH2O/s in different studies (4, 5, 6, 99, 121, 136). This is to be compared with a rate of 0.35–0.70 cmH2O/s in normal subjects subjected to airway occlusion during sleep (7–9). Although these rates cannot be directly compared because the lower lung volume in the obese OSA patients amplifies the relation between muscle activation and occlusion pressure (for review, see Ref. 132), the differences in ΔP/Δt are more than can be accounted for by such a mechanism.

Recently, the breath-by-breath increase in V̇E during nonrapid eye movement (NREM) sleep on CPAP was determined following step changes in inspired gases in 21 patients with severe OSA [AHI = 91 ± 24 h⁻¹ (129)]. This approach has the advantage of expressing the rate of change in chemical drive in the same units in which T ER and arousal threshold can be expressed (i.e., %increase, relative to eupneic drive, required to elicit either response). In this fashion the interrelation between different variables that contribute to instability can be examined in individual patients. By use of a range of inspired gas mixtures, it was possible to estimate the %increase in chemical drive following a moderately severe obstructive hypopnea or an apnea associated with a modest 4% reduction in O2 saturation (129). The average response is shown in Fig. 6 (heavy solid line) along with the range observed (dashed lines). As expected from the circulation delay, there was no increase in drive until breath 3. Beyond breath 3 the rate of increase was, on average, remarkably high [134 ± 77% of eupneic drive between breaths 3 and 5 (Fig. 6)]. As with all other variables that affect breathing stability in these patients, the rate of increase in chemical drive in the span of two breaths (breath 3 to breath 5 in this case) varied enormously among patients (49–380% of eupneic drive). Thus the impact of circulatory delay on postevent increase in chemical drive is highly variable and can be extremely large. Clearly, for any given patient, the rate of rise of chemical drive would be lower or much lower if the hypopnea were not as severe, or higher if desaturation was >4%.

To illustrate the interrelation between the various stability factors, the average arousal threshold (TA) and representative examples of T ER in these patients are also shown in Fig. 6. Arousal threshold ranged from 40% to >280% of eupneic drive (129). T ER in the same patients ranged from 0 to >174% of eupneic drive (129; Fig. 4). Two T ER values, one above (TER1) and one below (TER2) arousal threshold, are shown. Where T ER is higher than TA (e.g., TER1), whether because T ER is quite high, TA is quite low, or both, arousal is inevitable, regardless of the rate of rise of chemical drive. Because of the highly destabilizing influence of arousal, instability is very likely to occur in such cases unless the effect of arousal on flow is modest. Where T ER is less than TA (e.g., TER2), arousal may or may not occur, depending on the difference between the two thresholds (i.e., TA – T ER), and on circulation delay and the rate of rise in chemical drive, which determine how much chemical drive will rise during the delay time. In the example shown (Fig. 6, line B), arousal would still occur if circulation delay were as it is most commonly, two or more breaths but would not occur if circulation delay were only one breath or if the rate of rise of chemical drive were much below the average observed in these patients (e.g., Fig. 6, line C). The same variables also determine the temporal relation between airflow opening and arousal, if one occurs. Thus, with line A of Fig. 6, both reflex opening and arousal would occur very early in breath 3 (i.e., almost simultaneously, thereby obscuring the fact that the airway opened reflexly). With line B of Fig. 6, arousal would occur one breath after reflex opening, whereas with line C it would occur two or three breaths later, or may not occur at all. All these time relations between reflex opening and arousal were observed and documented (128). This plot, therefore, accounts for the near inevitability of arousals at the time of airflow opening in patients with recurrent obstructive events and illustrates why their occurrence need not indicate that arousal was required for opening to occur.

Without arousal [i.e., (TA – T ER) > (circulation delay × rate of increase in drive)] the overshoot is not as likely to be large (Fig. 5). However, a large overshoot may still occur if T ER is high, or if chemical drive increases substantially during the circulation delay, resulting in more complete airflow opening.

It is important to note that the various factors that contribute to an overshoot, and hence instability, show little or no correlation with each other (129). Thus the mechanisms contributing to instability vary considerably among patients (129). All the patients represented in Fig. 6 had continuous instability during sleep. Yet some had a low rate of rise of chemical drive, while others had low T ER and/or high arousal threshold (129). Accordingly, instability develops when a critical combination of these variables exists that results in a sufficiently large overshoot.

DIFFERENCES BETWEEN SUBJECTS WITH AND WITHOUT OSA

Without arousal [i.e., (TA – T ER) > (circulation delay × rate of increase in drive)] the overshoot is not as likely to be large (Fig. 5). However, a large overshoot may still occur if T ER is high, or if chemical drive increases substantially during the circulation delay, resulting in more complete airflow opening.

It is important to note that the various factors that contribute to an overshoot, and hence instability, show little or no correlation with each other (129). Thus the mechanisms contributing to instability vary considerably among patients (129). All the patients represented in Fig. 6 had continuous instability during sleep. Yet some had a low rate of rise of chemical drive, while others had low T ER and/or high arousal threshold (129). Accordingly, instability develops when a critical combination of these variables exists that results in a sufficiently large overshoot.

Without arousal [i.e., (TA – T ER) > (circulation delay × rate of increase in drive)] the overshoot is not as likely to be large (Fig. 5). However, a large overshoot may still occur if T ER is high, or if chemical drive increases substantially during the circulation delay, resulting in more complete airflow opening.

It is important to note that the various factors that contribute to an overshoot, and hence instability, show little or no correlation with each other (129). Thus the mechanisms contributing to instability vary considerably among patients (129). All the patients represented in Fig. 6 had continuous instability during sleep. Yet some had a low rate of rise of chemical drive, while others had low T ER and/or high arousal threshold (129). Accordingly, instability develops when a critical combination of these variables exists that results in a sufficiently large overshoot.
collapsible airways and at least some degree of OSA (127–129). It is clear that within this patient group there are major differences in the response to sleep-induced obstructions and that these differences largely account for the range of polysomnographic abnormalities observed (i.e., range of AHI). The important question arises, however, as to whether the range of responses in these patients is representative of the responses in the population at large. While more research needs to be done in this area, limited data suggest that the range of responses to upper airway obstruction is similar in OSA patients and in others with more favorable pharyngeal mechanics. Thus King et al. (56) found that when normal subjects are made to sleep with negative airway pressure they develop OSA that is indistinguishable from that observed with natural collapse in OSA patients. Notably, the AHI varied widely among subjects (from near zero to 90 h\(^{-1}\)), event duration was similar to that found in natural OSA (\(\approx 18\) s on average), and most but not all events were terminated by arousal (56). From the data presented, it is also clear that, as in patients with natural OSA (127), these subjects had periods of stable breathing at the same negative pressure that was associated with OSA at other times.

SUMMARY OF DETERMINANTS OF POLYSOMNOGRAPHIC FEATURES OF OBSTRUCTIVE SLEEP DISORDERS

Figure 7 is a flow chart that summarizes how different control mechanisms determine the response to the primary anatomic problem and, hence, the polysomnographic features. At the top of the chart, the anatomic defect is not expressed in absolute terms (i.e., \(P_{\text{Close}}\) or \(V_{\text{MAX}{0}}\)) but as the relation between the anatomic defect (quantified by \(V_{\text{MAX}{0}}\)) and the individual’s flow demand. This is to emphasize that the same anatomic defect may or not be significant depending on the metabolic and gas exchange status. Distinction is made between \(V_{\text{MAX}{0}}\) being higher or lower than 30% of peak flow demand. This is because when \(V_{\text{MAX}{0}}\) is \(>30\) peak unloaded flow (i.e., mild/moderate hypopnea at eupneic drive), compensation is possible without the need to increase \(V_{\text{MAX}}\), whereas with more severe obstructions at eupneic drive a stable response requires an increase in \(V_{\text{MAX}}\), which is a less reliable and potentially more unstable response. With milder obstructions (left side of Fig. 7 chart), the entire range of polysomnographic features, from no abnormality to steady flow limitation to OSA, may occur, depending on \(T_{\text{ER}}\) and the potency of the duty cycle response. With more severe hypopneas and apneas an equally wide range of manifestations may result, depending on the relation between \(T_{\text{ER}}\) and arousal threshold, absolute arousal threshold, and loop gain (which incorporates circulation delay, and dynamic plant and controller gains).

Regardless of the initial severity, the final responses may be normocapnic or hypercapnic, depending on the steady-state response to CO\(_2\) and hypoxia and on arousal threshold. Daytime hypercapnia may result particularly in the presence of comorbidities (e.g., chronic obstructive pulmonary disease [the overlap syndrome (114) or obesity-associated restrictive disorder [obesity hypoventilation syndrome; Ref. 78])]. A low \(T_{\text{ER}}\) mitigates hypercapnia even when CO\(_2\) and hypoxic responses are low since the change in gas tensions required to increase \(V_{\text{MAX}}\) to a sustainable level would still be small. This
accounts for the absence of a hypercapnic version of stable breathing on the far right of the chart in Fig. 7.

The AHI is the product of the frequency of cycling when breathing is unstable and the percentage of sleep time spent cycling (127). The frequency of cycling during unstable periods is principally determined by the rate of increase in chemical drive, and the lower of TER and arousal threshold. A low cycling frequency results when the rate of rise of chemical drive is low and both TER and arousal threshold are relatively high. A low rate of rise may be a reflection of 1) mild ventilatory deficit, resulting in a very slow deterioration of blood gas tensions. In such cases, hypopnea duration may be several minutes long and would be associated with mild blood gas tension changes, a picture currently identified as the upper airway resistance syndrome (34). 2) On the other hand, the ventilatory deficit may be severe but the response to the chemical changes (chemosensitivity) is depressed. Here, the hypopnea may again be quite long but would be associated with severe abnormalities in blood gas tensions. In both cases cycling frequency is low, but the mechanisms, mechanical abnormalities, and clinical consequences are clearly very different.

When either TER or arousal threshold is relatively low, or when rate of rise of chemical drive is high (more severe obstructions and/or high loop gain), such protracted (minutes) hypopneas/apneas are not possible and cycling frequency will tend to be high.

The percentage of time spent in instability, the other component of the AHI, is related to how far removed from stability the system is, relative to how much changes in the stability factors can occur spontaneously during the night. Some of the factors that determine stability change from time to time during sleep. Chief among these is arousal threshold (2). However, other factors may also change, although this needs experimental confirmation. For example, TER may be different in different body or head positions because of position-related differences in passive mechanics. Alternatively, passive mechanics may change through the night as a result of changes in vascular or surface tension properties of the pharynx. Lungs-carotid transit time may also change with cardiac function through the night. If the mechanism of instability is such that the within-night changes in stability factors cannot bring the system to a stable state, cycling will persist throughout sleep and the AHI will equal cycling frequency. For example, if TER is sufficiently higher than arousal threshold that the spontaneous changes in arousal threshold cannot place the two thresholds in a stable relationship, there cannot be stable time. On the other hand, if the constellation of stability factors is such that the spontaneous changes that occur during the night are enough to bring the system to a stable state, cycling will occur at times but not at others. For example, entering a deeper sleep state, or a change in body or head position may abort cycling, and vice versa.

Finally, the form an obstructive event takes is related to the severity of the mechanical abnormality and the extent of ventilatory overshoot. When passive Pclose is negative the obstructive event can only be a hypopnea. When passive Pclose is >0, the obstructive event may be either a hypopnea or an apnea, depending on whether the overshoot reduced chemical drive below TER. In either case, if the overshoot is sufficiently large to reduce chemical drive below the apneic threshold (101), the obstructive event may be preceded by a period of central apnea (i.e., mixed apnea). Two additional phenomena may help fashion the obstructive event although their actual contributions remain uncertain. First, the ventilatory overshoot following an obstructive event is a form of active hyperpnea. Such active hyperpneas do not disappear promptly as the stimulus that caused them is removed. A trailing central stimulation follows for a variable period [short-term potentiation (STP) or afterdischarge (23–25)]. A potent afterdischarge mechanism may mitigate the effect of washing away the chemical stimulus during the overshoot and maintain respiratory drive at a higher level in the overshoot’s aftermath than it would otherwise be (66, 125). Such an action may avert central apneas at the beginning of the events or result in an obstructive hypopnea where an obstructive apnea may have otherwise followed. The afterdischarge was found to be, on average, less potent in patients with severe OSA than in normal subjects (28, 51), suggesting that an attenuated afterdischarge mechanism may contribute to the development of mixed apneas in some patients. Second, the pharynx displays viscoelastic properties in that its elastic recoil (i.e., collapsibility; e.g., Pclose) is least following a sustained inflation (e.g., immediately after reduction of CPAP) and increases in a time-dependent manner following removal of the inflating force (98, 127). During the ventilatory phase following an obstructive event the pharynx is maintained open for a finite period. Depending on how strongly these viscoelastic properties are expressed, collapsibility will decrease to a variable extent during the ventilatory phase and increase again later as the pharynx narrows. Such time-dependent changes in passive mechanics may delay the onset of obstruction in the aftermath of the overshoot and may, as in the case of the afterdischarge mechanism, help convert some obstructive apneas into hypopneas.

**REFERENCES**


**GRANTS**

This review was supported by the Canadian Institutes of Health Research.

**DISCLOSURES**

M. Younes holds several patents on technology related to mechanical ventilation. Some of these have been licensed to ventilator companies, and Dr. Younes receives royalties from these licenses. None of these patents or licenses has any relation to the work or subject matter of this review article.
77. Okabe S, Chonan T, Hida W, Satoh M, Kikuchi Y, Takishima T. Role of chemical drive in recruiting upper airway and inspiratory inter-
MECHANISMS OF OBSTRUCTIVE APNEA


