# Embodiment, not imitation, leads to the replication of timing phenomena

Piers Messum

112 Warner Road, SE5 9HQ London, UK
p.messum@gmail.com

In many languages a spoken vowel is shorter before phonologically voiceless consonants than before voiced ones. In West Germanic languages, including English, tense and lax vowels have different lengths and there is a characteristic 'stress-timed' rhythm. For these phenomena and some others, it is generally assumed that time is the controlled variable for production, and that children replicate these speech behaviors through the reproduction of timing patterns abstracted from the adult input (i.e. by imitation).
The developmental data does not support these assumptions, and an imitative mechanism would present a young speaker with a highly complex challenge. Instead, these behaviors are more plausibly the result of the embodiment of speech. Embodiment goes beyond laryngeal and articulatory function. A child's speech breathing is not a scaled-down version of the adult model but a distinctly different skill, and one that must be learnt during speech production. Similarly, the aerodynamic setting of child speech differs significantly from that of adults. The constraints that these factors impose become manifest in speech as changes in timing, but these changes are epiphenomenal, not modelled directly. Phenomena particular to West Germanic languages reflect the style of speech breathing these languages require of a child.

# 1 Introduction

I will describe how the allocation of limited aerodynamic resource within an embodied system is a single mechanism that explains three apparently distinct temporal phenomena:

1. The shortening of the preceding vowel when a phonologically voiceless consonant closes a syllable (so that in English the vowel in *seat* is shorter than that in *seed*). This phenomenon goes under many names: here I will use 'pre-fortis clipping' (PFC). PFC is found in most languages where it can potentially occur.
2. The 'compression' of the vowel as segments are added to the coda of a syllable (e.g. [1]), as in *ram, ramp, ramped*. By analogy to PFC, I will call this pre-consonant-cluster clipping (PCCC).
3. In West Germanic languages, the progressive shortening of syllables as the number of syllables in a foot progressively increases, which I will refer to as foot level shortening (FLS). This is the most important contributor to the percept of 'stress-timing' although it does not lead to true isochrony of stressed syllables.

I propose that these effects appear in children as a consequence of the style of speech breathing that they must use, given the differences between their bodies and those of adults.

The mechanism is more fully described in Part 1 of [2], which also contains explanations for the conjunction of properties that distinguish tense and lax vowels in English, for patterns of VOT, and for some other phenomena that have proved problematic in phonetics. In [2] I argue that none of these are replicated by imitation, and that children do not learn how to pronounce speech sounds by imitation, either. (Part 2 describes how children do this instead through the mirroring activity of their caregivers.) I will refer to sections in [2] as Mx.x.

# 2 Explaining temporal phenomena

Temporal changes in the acoustic signal must reflect changes in the movements of the upper articulators. These movement changes need not always be the result of modifications to high-level motor commands; in some cases they may occur for purely mechanical reasons, with control at a detailed level forgone by the speaker. However, if high-level motor commands are being modified this can result from at least two distinct motivations:

1. The need to satisfy a physiological/mechanical/aerodynamic or other constraint. (The change in articulator movements necessary to achieve this lead to timing effects in the output, but these are just a by-product of an essentially non-temporal process.)
2. An 'intention' to alter the timing *per se*.

In the first case, the mechanism by which modifications are planned would be independent of the speaker's perceptual criteria. Thus while he might make use of a given timing phenomenon to support his perception of others' speech, this would not inform his own production.

In the second case, the planning mechanism would operate through a set of linguistic/phonological 'rules' derived from the speaker's perception of ambient speech. So during the period of acquisition he would have abstracted and modelled the regularities that he noticed in the speech of others, and he would now be using the model he created to control his production. From a developmental perspective it would be natural to describe this as a form of imitation.

(The reason why such 'rules' might have appeared in a language in the first place has been controversial and is not my concern here. However, a common assumption has been that if they create some perceptual benefit then this will have led to them becoming embedded in the grammar.)

Let us label any mechanism of the first type 'phonetic' and any of the second 'phonological', based on where we believe that the decisive influence on the articulators derives from in each case (the embodiment of the speech apparatus or the mental grammar of speech, respectively). I have presented the mechanisms as if they are alternatives, but one may be dominant at one stage in development and then be succeeded by the other; and there is the further possibility of overlap at some times.

For the three 'timing' phenomena being considered here, there has been no success in finding satisfactory phonetic mechanisms; see, for example, [3, 4, 5] for reviews of attempts to explain PFC this way.

However, the internal workings of phonological mechanisms have proved no easier to explain. If there is any reality to the putative rules that determine the timing patterns they produce it should be possible to model their operation, but [6], for example, failed to do this for PFC. In fact, this has almost always been the outcome of attempts to model 'timing' phenomena of this kind. Most recently, for example, [7] failed to find any straightforward explanation for their data on boundary related lengthening.

I will describe a phonetic mechanism that differs from previous proposals in its consideration of what has been the 'Cinderella' of speech research: speech breathing.

# 3 Speech breathing in children

Let us define speech breathing as the co-ordinated actions of (1) the muscles that change lung and airway volume with (2) the muscles that have a valving effect on the vocal tract, in order to produce the aerodynamic conditions required for speech. Speech breathing is a complex motor skill not only because of the large number of muscles involved in both (1) and (2), but also because one contributor to the alveolar (lung) air pressure ($P_{alv}$) is the pressure generated from distended body tissue (as in a balloon with an elastic rubber skin). As the volume of the lungs changes continuously while we speak, so does this non-volitional contribution to $P_{alv}$, called either 'relaxation' or 'recoil' pressure. With this mechanical backdrop, there is a need for constant adjustment of the volitional contribution made by the chest wall musculature if subglottal pressure ($P_{sg}$) is required to remain constant.

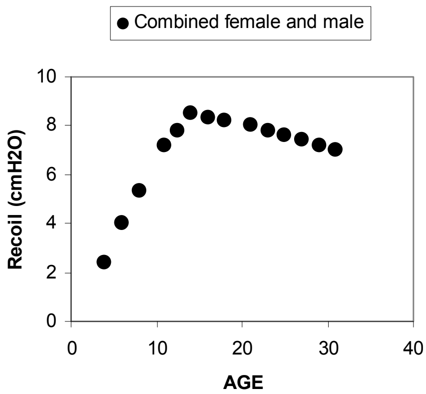


Fig. 1. Redrawn from [8] figure 7. Original caption: "Static recoil pressure".

Figure 1 shows the relaxation (recoil) pressure generated by speakers of various ages at the end of a normal inspiration. Children's relaxation pressures are lower than those of adults because the compliance of their lung and chest wall tissue is greater [9]. (Children's systems are 'floppier'.)

Adults typically speak with $P_{sg}$ of 6-8 $cmH_2O$, children on higher pressures. So the adult style of speech breathing – inflating the lungs and then largely speaking on relaxation pressure (upper graph of Figure 2) – is not available to young children.

A simple way of thinking about this developmental difference is to imagine that when an adult inhales he is inflating a balloon inside him which then provides a reservoir of already pressurised air for speech. This need only be supplemented by volitional expiratory activity as the balloon gradually deflates. When a child inhales he is inflating not a balloon but a paper bag, and thus he has, to a first approximation, only a reservoir of unpressurised air as a result. He must create the pressure he needs with significant volitional activity prior to and throughout the course of speech. (Lower graph of Figure 2.)

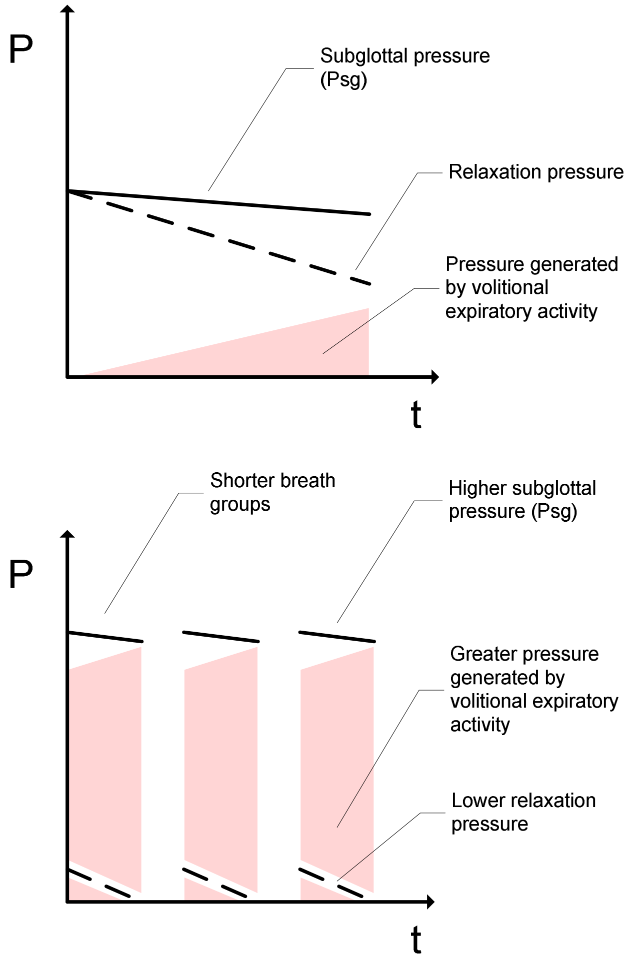


Fig. 2. Schematic illustration of the relative importance of volitional expiratory activity in an adult (above) and child (below).

Over the course of an utterance, adults normally develop a relatively constant $P_{sg}$: an 'elevated background pressure' that is generated within the chest wall, and valved by the actions of the upper articulators. This aerodynamic profile is facilitated by the relaxation pressures they develop and is compatible with the smooth respiratory drive that is apparently used. However, [10] demonstrated that under certain conditions even the adult style of speech breathing becomes pulsatile; when, for example, the rate of articulation slows (see Figure 3) or when high flow (high effort) segments are produced. (He also drew attention to the articulatory skill required to maintain smooth speech breathing.)

Note that a pulsatile style of speech breathing would normally involve a series of net-expiratory gestures within a single breath group. This is not illustrated in the graphs of Figure 2.

A child speaks

- with negligible contribution from relaxation pressure;
- at a slower rate than an adult;
- with all segments being effectively 'high effort' (from the point of view of the respiratory system);
- at a time when articulatory skill is still being developed.

For these reasons (discussed in detail in M3.4, together with some others), we should expect a child's respiratory drive to be not smooth but pulsatile; initially with one pulse for each syllable produced.

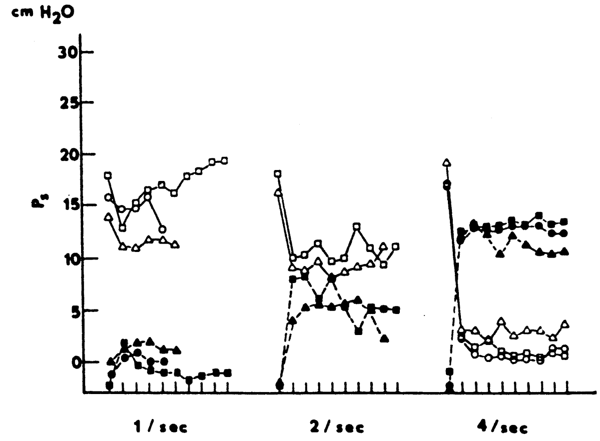


Fig. 3. As rate increases, there is a move from discrete changes in background pressure to an essentially smooth contour. (Unfilled symbols are the change in $P_{sg}$; filled symbols are the background level of $P_{sg}$.)
Figure 37 from [10]. Original caption: "Subglottal pressure data: subject TK repeating /s/ at fixed rates."

The variable he uses for control of these pulses is most plausibly the percept of the effort he has applied with his respiratory system musculature. (It might alternatively - or additionally - be some measure of the $P_{sg}$ initially generated by a pulse of expiratory activity without affecting the argument that follows.) The quantum or level of this effort is unlikely to be determined by what in an adult we would consider to be the segmental content of the syllable produced, which during the early stages of speech will be mostly CV, VC and CVC forms. So while the effort applied for a syllable will vary depending on, for example, the overall loudness that is to be achieved, it will not vary according to the different 'segments' within the domain of a pulse. The action of the respiratory system and the upper articulators will be in a 'frame and content' relationship.

# 4 A phonetic (embodiment) mechanism for PFC

This model of SB in children allows a straightforward explanation for PFC, which starts with the recognition that fortis and lenis consonants differ in their significance for the respiratory system. For while there has been some debate and disagreement about whether fortis consonants require more 'articulatory force' in production than lenis ones, there is no doubt that they require more aerodynamic resource. Compare, for example, the respiratory system effort required to produce an [s] (with only a turbulent sound source), with that required for a [z] (where the efficient sound producing mechanism of vocal fold vibration is additionally employed). Attempting to prolong each sound indefinitely makes the disparity very clear.

Thus in a system where (1) there is the same quantum of aerodynamic resource available to produce both of, say, *peace* and *peas*, and where (2) greater resource is required for the final (fortis) segment in the former, then this resource must be allocated at the expense of the other segments; in particular at the expense of the preceding vowel. Thus we see 'clipping' of the vowel not because the speaker intends to reduce its duration *per se*, but because he must plan to produce it with less resource (more quickly) if the aerodynamic demands of the syllable are to be met.

# 5 PFC: phonetic or phonological?

We can now examine some of the characteristics of PFC to see whether the phonetic mechanism described or a phonological (imitative) one provides a more plausible explanation for the phenomenon. The following points all tell in favour of the phonetic account (and I am not aware of any significant points that would argue against it).

## 5.1 Contextual dependence

*In adult speech, PFC is evident in primary stressed syllables, less evident in secondary stressed syllables and negligible or absent in unstressed syllables [11]. It is clear in citation forms, but often absent in normal speech [12].*
In the phonetic account, both of these findings follow naturally from the 'clipping' mechanism being dependent on respiratory drive that is applied in a pulsatile, quantised fashion. In M3.3 I argue that the style of speech breathing used by a mature speaker has the potential to vary along a continuum, with an adult's speech breathing likely to be most pulsatile in exactly the contexts where PFC is most evident, and not pulsatile (i.e. smooth) in contexts where PFC is absent. (This continuum approach also allows us to reconcile some apparently contradictory results on styles of speech breathing in the literature.)
On the other hand, it would seem curious if these patterns of context dependency were the result of phonological rules, particularly if the rules were supposed to make the identification of syllable-final consonants easier. They would rather oddly be making no contribution to this identification in those contexts where a consonant's articulation would be least clear.

## 5.2 Extent of phenomenon

*PFC is comparable in French and English [13].*
Previously PFC had been thought to be more marked in English, e.g. [14], implying that it has a phonological basis. This conclusion can no longer be drawn from the data.
*PFC shortens the /n/ of "tent", and the /l/ of "milk" [15].*
This is a natural result of the phonetic mechanism; it is not clear that it is a natural result of a putative phonological rule set.
*English speaker vowel length ratios for PFC vary from around 1.2 to 1.8 [16].*
Again, such variability is to be expected with the phonetic mechanism, since any timing effects it produces are only by-products of a process which has no temporal targets. However, the range of shortening seen is surely surprising if PFC is learnt by imitation and if timing is the variable being directly modelled by speakers.

## 5.3 Occurrence in non-standard speech

*PFC occurs in whisper [17].*
In the past this has been taken as evidence that the phenomenon is phonological rather than phonetic.

However, [18] and others have shown that the glottis narrows in whispered lenis consonants compared to fortis ones, so the aerodynamics of the two situations are not neutralised. A narrowed glottis would require less aerodynamic resource to create a turbulent sound source.

*PFC depends upon the phonological status of the consonant, rather than actual vocal fold vibration [19].*

[20] has shown that even if a final /z/ is devoiced its aerodynamics are similar to a voiced /z/ rather than to an /s/, and PFC occurs with both types of /z/. (There is discussion of the aerodynamics of other consonants pairs in M4.4.)

## 5.4 Evidence from 'universality'

*PFC is almost a language universal (in languages allowing syllable codas that have not been neutralised with respect to voicing). Czech, Swedish, Saudi Arabic and Polish are the known exceptions [21, 22, 23].*

The first three of these languages use vowel length phonemically. For Swedish, [21] report 24-month-old children exhibiting PFC before suppressing it 6 months later (in order, presumably, to properly express phonemic vowel length).

This data is totally inconsistent with any phonological account of PFC since Swedish children would never hear the phenomenon in the speech of others and thus would have no model to imitate.

## 5.5 Development

*PFC appears in the speech of even very small children (perhaps from as early as 20 months [24])*

In M7.3 I considered the imitative acquisition of 'timing' phenomena including PFC from various perspectives:

- young children's motivation;
- young children's capacity for the modelling that is supposed to occur;
- the data from developmental studies.

From each viewpoint, problems exist with any account of replication based on a child developing phonological rules through some imitative process.

# 6 A phonetic mechanism for other 'timing' phenomena

The mechanism described for PFC also explains PCCC, the 'compression' of the vowel as segments are added to the coda of a syllable. For example, as in *ram, ramp, ramped*. As more elements are added for production with an unvarying pulse of respiratory system effort, so the resource allocated to each must be diminished and the time of individual execution reduced.

The mechanism then extends naturally to also explain foot level shortening (FLS), a phenomenon in West Germanic languages that contributes significantly to the percept of 'stress-timing'. In fact, FLS has never been explained without an appeal to rhythmicity as its motivation, and it is therefore taken as strong evidence by those who believe that speakers plan the production of stressed syllables with something like a 'tendency to isochrony'.

To develop an alternative account of FLS, we need to first examine the nature of stress-accent in English, German, Dutch etc, and then the aerodynamic characteristics of the syllables that make up prosodic feet.

Stress-accent languages create prominence by making stressed syllables louder as well as longer [25]. There is a continuing debate among phoneticians about whether or not this requires an adult speaker to increase his respiratory drive for routine sentence stress [e.g. 26, 27]. However, at around 2 years of age (when some children start to deploy stress-accent) and for some years afterwards there seem very good reasons to believe that children learning West Germanic languages will increase their respiratory system drive on stressed syllables. See M5.2 for discussion.

At the same time, these children are starting to change the way that they produce unstressed syllables, including the weak syllables that may follow a stressed syllable in a foot. The vowel reduction that is involved creates segments that are aerodynamically very different from normal vowels. [28] reports a comparison between reduced and full vowels in comparable CVC contexts where durations had means of 30 ms and 110 ms respectively, and a typical opening of the vocal tract might be 0.2 $cm^2$ as compared to 2 $cm^2$.

Thus the aerodynamic resistance offered by the 'vocalic' portion of a weak syllable is determined by two characteristics: the change between the consonantal states is (1) very brief and (2) minimal in its extent. From the perspective of the respiratory system, these weak syllables are not dissimilar to consonant clusters and syllabic consonants. To a first approximation, all create periods of high resistance to airflow.

So where a foot contains a stressed syllable followed by zero or more reduced syllables, the respiratory system sees the coda of the stressed syllable plus all of the following reduced syllables as a single high resistance unit; in a sense, as a complex consonant cluster. *Fricatives and resonants* would appear as || CCVC·C·CC·C | CVC·C·CCC ||.

We can now see how the mechanism that explained PFC and PCCC would also operate at the level of the foot in West Germanic languages, rather than just at the level of the syllable as in other languages. FLS occurs in languages that demand a pulsatile style of SB in a young speaker in order to create stress prominence. He then has to distribute an invariant amount of aerodynamic resource over the domain of a stressed syllable plus following unstressed ones. The addition of extra weak syllables takes resource away from existing ones, which therefore have to be articulated more quickly.

# 7 Summary

I have described a unifying explanation for PFC, PCCC and FLS in which all three emerge as a result of the need to distribute a limited aerodynamic resource among parts of a syllable or foot that have different aerodynamic requirements. The model of speech breathing that supports this also accounts for accentual lengthening, phrase-final lengthening and P-centres. (See M6.4)

The consideration of further aspects of speech breathing in children and the aerodynamics of speech in a child-sized body generates explanations for the conjunction of properties of tense and lax vowels in West Germanic languages, for the patterning of voice onset time data, for declination and for some other phenomena. These are all described in [2].

It is satisfying that an apparently arbitrary and disparate set of 'timing' and other phenomena in English can be given principled explanations which bind them into a now more coherent whole.

The mechanisms I propose do not suffer from the shortcomings of phonological accounts with respect to acquisition. They do not require us to believe that a young child becomes a junior phonetician at a time when he has so many competing and more important demands on his attention. He must, of course, become very familiar with his own production system, but he need not study speech phenomena that are not linguistically distinctive.

However, some of the explanations in [2] rely on the differences of scale between the production apparatus in children and adults. Why, then, do the phenomena in question appear in adult speech when those constraints that only operate on children no longer apply?

Among various possibilities for this, I explain in M8.2 why I prefer an account where, to a first approximation, adult speakers plan speech in a way that is decoupled from how they produce it. So with respect to the changes in loudness associated with stress (for example), speakers plan the articulation of a foot so that the movements of the upper articulators can be executed with any style of speech breathing, including a highly pulsatile one. The actual style used is 'decided' after this planning stage, based on a variety of factors. The timing of English speech is then coherent across whisper, normal speech, shouting etc.

# References


[1] K.Munhall, C.Fowler, S.Hawkins, and E.Saltzman, "'Compensatory shortening' in monosyllables of spoken English", *J.Phon* 20, 225-239 (1992)

[2] P.R.Messum, "The role of imitation in learning to pronounce." PhD thesis, University College London (2007) (http://p.messum.googlepages.com/home)

[3] L.Lisker, "On 'explaining' vowel duration", *Glossa* 8, 233-246 (1974)

[4] K.R.Kluender, R.L.Diehl, and B.A.Wright, "Vowel length differences before voiced and voiceless consonants: an auditory explanation", *J.Phon* 16, 153-169 (1988)

[5] N.Braunschweiler, "Integrated cues of voicing and vowel length in German: a production study", *Language and Speech* 40, 353-376 (1997)

[6] H.S.Gopal, "Generalizability of Current Models of Vowel Duration", *Phonetica* 53, 1-32 (1996)

[7] A.E.Turk and S.Shattuck-Hufnagel, "Multiple targets of phrase-final lengthening in American English words", *J.Phon* 35, 445-472 (2007)

[8] E.T.Stathopoulos, "A review of the development of the child voice: an anatomical and functional perspective", in: *Child Voice*, P.White, ed., KTH, Stockholm (2000)

[9] N.K.Russell and E.Stathopoulos, "Lung volume changes in children and adults during speech production", *JSHR* 31, 146-155 (1988)

[10] T.R.Kneil, "Subglottal pressures in relation to chest wall movement during selected samples of speech." PhD thesis, University of Iowa (1972)

[11] K.J.de Jong, "An articulatory study of vowel duration changes in English", *Phonetica* 48, 1-18 (1991)

[12] T.H.Crystal and A.S.House, "The duration of American-English vowels: an overview", *J.Phon* 16, 263-284 (1988)

[13] C.Laeufer, "Patterns of voicing-conditioned vowel duration in French and English", *J.Phon* 20, 411-440 (1992)

[14] M.Chen, "Vowel length variation as a function of the voicing of the consonant environment", *Phonetica* 22, 129-159 (1970)

[15] J.C.Wells, *Longman Pronunciation Dictionary*, Longman, Harlow (1990)

[16] B.L.Smith, A.R.Bradlow, and T.Bent, "Production and perception of temporal contrasts in foreign-accented English", in: *15th ICPhS*, M.J.Solé, D.Recasens, and J.Romero, eds., Causal Productions, Barcelona (2003)

[17] D.Sharf, "Vowel duration in whispered and in normal speech", *Language and Speech* 7, 89-97 (1964)

[18] S.L.Hamlet, "Vocal fold articulatory activity during whispered speech", *Archs Otolar.* 95, 211-312 (1972)

[19] T.Walsh and F.Parker, "Vowel length and 'voicing' in a following consonant", *J.Phon* 9, 305-308 (1981)

[20] C.L.Smith, "The devoicing of /z/ in American English: effects of local and prosodic context", *J.Phon* 25, 471-500 (1997)

[21] E.H.Buder and C.Stoel-Gammon, "American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing", *JASA* 111, 1854-1864 (2002)

[22] F.M.Mitleb, "Voicing effect on vowel duration is not an absolute universal", *J.Phon* 12, 23-27 (1984)

[23] P.Keating, "Universal phonetics and the organisation of grammars", *UCLA Working Papers in Phonetics* 59, 35-49 (1984)

[24] M.A.Naeser, "The American Child's Acquisition of Differential Vowel Duration", W R & D Center for Cognitive Learning, Technical Report No. 144, University of Wisconsin (1970)

[25] M.E.Beckman, *Stress and Non-Stress Accent*, Foris (Netherlands Phonetics Arch. No. 7), Dordrecht (1986)

[26] J.Ohala, "Respiratory activity in speech", in: *Speech Production and Speech Modelling*, W.J.Hardcastle and A.Marchal, eds., Kluwer Academic Publ., London (1990)

[27] E.M.Finnegan, E.S.Luschei, and H.T.Hoffman, "Estimation of alveolar pressure during speech using direct measures of tracheal pressure", *JSLHR* 42, 1136-1147 (1999)

[28] J.C.Catford, "'Rest' and 'open transition' in a systemic phonology of English", in: *Systemic Perspectives on Discourse Vol 1* , W.S.Greaves and J.D.Benson, eds., Ablex, Normal NJ (1985)