Abstract: This paper examines, analyzes, and synthesizes literature on VOT production and perception in the context of Motor Theory of speech perception. Specifically, positive and negative evidences from speech production and perception, language acquisition, and evidence of dissociation from speech disorders of VOT are reviewed in the light of the Necessity Condition versus a weaker version of motor theory suggested by Hickok and Poeppel (2004), which postulates that the mapping of sensory representations of speech onto motor representations may not be an automatic consequence of speech perception, and indeed are not necessary for auditory comprehension (Hickok & Poeppel 2000), but can be activated and utilized strategically to assist in task performance.

I. General Theory of Speech Production and Perception
   a. Motor theory of speech perception
   b. Evidence from normal adults
   c. Developmental evidence

II. Production and Perception of Voice-onset Time (VOT) in Normal Individuals
   a. Acquisition of VOT
   b. Production and perception of VOT: categorical perception

III. Production and Perception of Voice-onset Time (VOT) in Individuals with Speech Disorders
   a. Overview of neural correlates of speech production and perception and the types of aphasia
   b. Nature of VOT production errors in Broca's aphasia and Wernicke's aphasia
   c. Perception of VOT in aphasic patients
   d. Evidence for motor theory of VOT perception
   e. Evidence for dissociation between VOT production and perception
   f. Other Evidence for dissociation of production and perception

IV. Conclusion
I. General Theory of Speech Production and Perception

a. Motor theory of speech perception

The Motor Theory of speech perception (Liberman et al., 1967, Liberman and Mattingly 1985) was proposed to address the problem of invariance perception despite highly varied acoustic input: that the acoustic patterns of the synthetic speech had to be modified if an invariant phonetic percept was to be produced across different context (Cooper et al 1952). The Motor Theory also addressed the question of what is the basic unit of speech perception, given the difficulty of finding invariance in speech acoustic unit that can map to a fixed mental representation. For example, the acoustic cues for identifying a stop consonant (e.g., /d/) vary depending on the transitional formant pattern into the vowel that follows (Figure 1). Therefore, in this case of /d/, it is impossible from a purely acoustic perspective to capture the unique characteristic formant patterns that defines /d/ and distinguishes /d/ from /b/.

Another example of such problems is the categorical perception—poor discrimination within categories (e.g., in a continuum from /da/ to /ga/) results not from degrees of differentiation in acoustic signals but from the fact that they can be produced with the same articulatory gestures. This mapping problem is also existent in the developmental domain, relevant to the issues of how children came to successful map varied speech signal onto the invariance of mental representations, and how is continuous speech signal being segmented and acquired (or is segmentation to the phoneme level necessary? See Morais et al 1994) in the first place. Indeed, the problem of one-to-one mapping from acoustic signal to phoneme categories is also demonstrated in the challenge of computational modeling of speech recognition and automatic speech phoneme segmentation algorithms.

In the light of these issues, the Motor Theory posits that we must abandon the idea of acoustic invariance for phoneme identification, and instead hypothesize that the invariance unit of perception is the articulatory gestures. While there are many revisions and variants of the theory over the years, a stronger version of Motor Theory of speech perception postulates that speech production is required (i.e., a prerequisite) for the successful decoding in speech perception, and that the perceived units of mapping variant speech signal to a invariance of mental representation involves the mapping of production onto perception, i.e., we hear the articulatory gestures from the speech sounds and decode from the gesture, instead of the acoustic signal itself. In the revised version of the theory (Liberman et al 1985), the proposed unit of perception is the speaker’s intended gesture (since the actual articulatory gestures are not arguably less varied than the speech acoustic signal itself), and that identification of the intended gesture is synonymous with identifying a loosely-defined ‘phonetic category’ whose relationship to phoneme or phone is not clear (Hawkins 2004). Therefore, as we will discuss later, any objections to the Motor Theory of speech perception must also consider alternative explanations to address these problems that it set out to solve.
b. Evidence from normal adults

(1) Mirror neurons and McGurk Effect

Despite its effectiveness in addressing the mapping problem described above, it has been pointed out that a strong version of the motor theory of speech perception is hard to maintain due to the fact such as our ability to perceive synthesized speech sound that are produced with no articulatory reality (e.g., sine-wave speech), and the dissociations between production and perception among aphasic patients (Zsiga 2011). The discovery of mirror neurons in the monkey motor system, which are claimed by some to support action understanding (Rizzolatti & Craighero, 2004), has led to a resurrection of the motor theory of speech perception (Fadiga & Craighero, 2006), including a strong version of the theory. For example, Fadiga and Craighero write, “Liberman’s intuition... that the ultimate constituents of speech are not sounds but articulatory gestures... seems to us a good way to consider speech processing in the more general context of action recognition” (Fadiga & Craighero, 2006, p. 489). In general, the mirror neurons refer to the observed activation of the brain region responsible for the production/action associated with the stimuli while perceiving the stimuli that is presented (perception linked to production). Vice versa, in the visual domain, single-unit recordings in the parietal lobe of primates have shown that many cells are sensitive not only to visual stimulation, but also to the monkey’s action towards that visual stimulation. For example, a unit (of processing) may respond not only when an object is presented, but also when the monkey reaches for that object in an appropriate way, even if the object is no longer in view (Murata, Gallese, Kaseda, & Sakata, 1996) (linking production/action to perception). Following the experiments that observed mirror neurons in a simple perception task where the listeners focused on the acoustic properties of the stimuli, Hickok and Poeppel (2004) proposed that the cortical processing system, upon receiving speech signals that are processed in the earlier stages in the auditory cortex (in the left Superior Temporal Gyrus, or STG region, responsible for speech perception), then diverges into two broad processing streams, a ventral stream, which is involved in mapping sound onto meaning, and a dorsal stream, which is involved in mapping sound onto articulatory-based representations.

The fact that mirror neurons of speech perception are only activated when speaker is focusing on the phonetic aspect (but not when then are in general focusing on the lexical-semantic aspect of the speech) is explained by this framework of two alternative processing mechanisms (Figure 2). It is also supported by the fact that the neural
processing mechanisms can be non-identical even though the stimuli presented remained constant (such as the speech-to-song illusion, see Zhang 2010). In addition to the mirror neuron effect of speech perception, the illusion of McGurk Effect also supports the Motor Theory of speech perception: when simultaneously presented with visual signal that contains information about articulatory gestures that are inconsistent with the acoustic information (e.g., visually /fa/ with acoustic /ba/), listeners perceive the speech sound that either showed an articulatory-dominance (/fa/) or a compromise between the two. This is consistent with the Motor Theory of speech perception: when both articulatory and acoustic information are salient and available, the brain utilize articulatory information to override the acoustic signal.

Consistent with the motor theory, recent functional imaging and transcranial magnetic stimulation (TMS) work has found that the motor speech system is indeed activated during speech listening (Watkins, Strafella, & Paus, 2003) and TMS interference of motor cortex has been reported to modulate speech perception under some conditions (Watkins & Paus, 2004). However, none of these studies provide a conclusive assessment of the causal role of the motor speech system in speech perception (Hickok, 2009, 2010).

(2) Perception in Production: Neural Feedback in Online Control

The significance of perceptual feedback in the speech production processes is well demonstrated in language acquisition literature as well as speech production/perception literature. It is well known that infants depend on auditory feedback to adjust their speech production, and such a production-perception linking mechanism is important for early success in speech production.

There has been accumulating evidence in recent years regarding the extent that speakers unconsciously utilize perceptual feedback to adjust their speech production (e.g., on-line control of the F0 production or loudness) (Chen et al 2007). In such experimental paradigms, subjects typically produce speech streams while almost simultaneously (usually with a very small amount of delay) receiving a gradually shifted feedback (e.g., shifted F0 feedback) through a headphone. It is observed repeatedly that under such condition, the speakers unconsciously adjusted their production (e.g., of F0) in the opposite direction of the perturbation to compensate for the altered perceptual feedback.
c. Developmental evidence

(1) Mirror neurons and activation of Broca's area in baby's speech perception

One specific question regarding the links between perception and production is how innate these links are. Liberman & Mattingly (1985) view the perception-action link for speech as potentially innate, whereas Kuhl & Meltzoff (1982, 1996) view it as forged early in development through experience. Imada et al. (2006) used magnetoencephalography (MEG) to study newborns, 6-month-old infants, and 12-month-old infants while they listened to nonspeech, harmonics, and syllables (Figure 3). Dehaene-Lambertz et al. (2006) used fMRI to scan three-month-old infants while they listened to sentences. Both studies showed activation in brain areas responsible for speech production (the inferior frontal, Broca’s area) in response to auditorally presented speech. Imada et al. reported synchronized activation in response to speech in auditory and motor areas at 6 and 12 months, and Dehaene-Lambertz et al. reported activation in motor speech areas in response to sentences in three-montholds. To determine the degree of innateness of the activation, newborns are tested by Imada et al. and showed no activation in motor speech areas for any signals, whereas auditory areas responded robustly to all signals. To sum up, these experiments suggest that perception-production linkages for speech develop by three months of age as infants produce vowel-like sounds (Kuhl 2008), consistent with the theory of an activated motor mechanism during speech perception in normal adults.

(2) Other evidences

Language acquisition literature showed mixed evidence regarding the motor theory of speech perception among
developing infants. On one hand, it has been demonstrated that babies also showed a McGurk Effect-like mechanism in their perception of the categorical continuum: when the same articulatory gesture image accompanies the acoustic signal, the infants cannot perceive a categorial distinction from /da/ to /ga/ (Kuhl et al 2008).

On the other hand, it has been suggested that the fact that babies can perceive and discriminate speech signal long before they can produce utterances is a counter evidence towards the motor theory of speech perception. However, as we have discussed above, even the three-month-old infants showed activation of speech production neurons in the Broca's area during speech perception, and it is clear that they have very little production at that time. Further experiments are needed to determine the degree of experience required for these mirror neurons among young infants, and the mechanisms that caused the delay from the time that the speech production neurons are activated during perception (early) and the time when they can actually produce the speech (later). Nonetheless, the early activation of the speech production neurons among three-month-olds demonstrates that the links of motor basis of perception can operate on a much lower level of neuro-cognitive processing than speech, a mechanism that is not domain-specific to speech processing and can be activated even by meaningless babbling.

d. Summary

To sum up, current evidence from normal individuals suggest that the links between production and perception is supported neuro-cognitively, at least in part. The question should be to what degree the Motor Theory of speech perception can be attested, to what degree this link is innate and domain-specific to language, and whether we can go as far as to say that speech production is a pre-requisite for successful speech perception and decoding, i.e., a strong version of the motor theory (here I term it the Necessity Condition). For instance, Hickok and Poeppel (2004) proposed that the mapping of sensory representations of speech onto motor representations may not be an automatic consequence of speech perception, and indeed are not necessary for auditory comprehension (Hickok & Poeppel, 2000). Instead, they postulate that motor representations of speech can be activated and utilized strategically to assist in task performance. In the following sections I focus specifically on the production and perception of VOT and address the motor theory of speech perception.

II. Production and Perception of Voice-onset Time (VOT) in Normal Individuals

a. Acquisition of VOT

Different languages usually utilize different sets of VOT contrast for categorical discrimination of voicing, for example, short lag/long lag (English) vs. pre-voiced/short lag (French). Also, VOT depends on place of articulation, with labial VOTs being shorter than velar and alveolar and, sometimes, alveolar being shorter than velar (Whalen et al. 2006). Here I examine literature on the perception (discrimination) and production of VOT among young infants and children.

1 This is a characterization of the Motor Theory of speech perception that I find the most convincing, consider the neuro-cognitive and anatomical evidence.
Cross-linguistic studies of infants' perception of VOT have found that infants from around 6 to 12 months of age are able to distinguish VOT contrasts not present in their native language environment (Lasky et al 1975; Aslin et al 1981), a finding that shows delayed language-specific discrimination tendency that are found in other types of phonetic perceptions, such as segmental contrast of vowel and consonants (Boysson-Bardies et al. 1984). For instance, Lasky et al (1981) found that among 4.5-6-month-old infants in a Spanish environment, subjects in the [−20 to +20 ms] group showed no signs of being able to discriminate these stimuli while the subjects in the [−60 to −20 ms] group and [+20 and +60 ms] group did. These results correspond to those reported for infants from English speaking families but do not correspond to those reported for Spanish speaking adults. Thus, the failure to find evidence of discrimination in the [−20 to +20 ms] group suggests that experience has little effect in determining bilabial stop phonetic categories in 4–6.5 mo-old infants (although it is unclear whether the reason for the failure of discrimination for the [-20 to +20 ms] group has a phonetic/perceptual grounding). Similarly, Aslin et al (1981) studied 6-12-month-old infants for discrimination along a synthetic VOT continuum ranging from -70 msec to +70 msec. Infants from an English-speaking environment provided reliable within-subject evidence for discrimination of VOT contrasts located at both the plus and minus regions of the VOT continuum. In addition, the adults showed heightened sensitivity to VOT differences near the English voiced-voiceless boundary in the plus region of the VOT continuum, a finding that was not evident in the infants' data.

Speech production data showed that the infants' production of VOT also corresponds to their perception in terms of the variability of the VOT produced in early years (at least around 1 year of age) that do not particularly show a strong native language-specific influence. Whalen et al (2006) looked at the babbling data of 9 and 12 month age infants from English and French environment, respectively. The results showed that there was little or no difference between the languages for duration of positive VOTs, which were usually in the “short-lag” range. The duration of prevoicing also did not differ between languages, but the proportion of prevoiced utterances did (French-learning infants: 44.2% prevoicing; English-learning: 14.3%). Labial, alveolar and velar stops differed in VOT, with alveolar longer than labial and velar longer than alveolar, which the authors attributed to a possible mechanical cause. The authors concluded that the lack of long-lag VOT indicates that the English-learning infants have not mastered aspiration by 12 months. In contrast, the different proportions of prevoicing, however, suggest that the French-learning infants attempt to imitate the prevoicing that is used frequently (and contrastively) in their native language environment. This difference was attributed to the different levels of difficulties in mastering the articulatory gestures of prevoicing and aspiration.

Whiteside et al (2003) demonstrated the decrease of variability in VOT production between age 5 to 13, with the youngest children of the 46 studied produced most variant values of VOT. The developmental patterns of variability in the VOT data are interpreted as evidence for increased levels of stability in speech output as a function of the maturing motor speech skills. The lack of significant age differences between the 11- and 13- year-olds suggests that levels of variability in VOT may begin to level off at this stage within the human lifespan.

To sum up, the data on the development and acquisition of language-specific VOT perception and production are consistent with each other, demonstrating a pattern of considerable variability in speech production and a non-language-specific ability of discrimination of VOT values beyond one year old of age. This is also consistent with the
motor theory of speech perception that postulates a good correspondence between the developmental stages of production and perception of VOT.

b. Production and perception of VOT: categorical perception

It has been well established that the perception of voicing categories and VOT in normal adults varies depending on the categorial boundaries of specific languages. Therefore, in a discrimination task, poor discrimination is usually expected with VOT values far away from the categorical boundaries, whereas the rate for correct discrimination peaks around the categorical boundaries. The location of the category boundary was found to differ by language experience. Therefore, for instance, for Spanish speakers, the [b] vs. [p] boundary fell at 15 ms positive VOT, whereas for Thai speakers [b] was identified for any VOT less than -20 ms, [p] for VOT between -20 and +40 ms, and [ph] for VOT above +40 ms. A comparison of VOT boundaries among English, Thai and Spanish speakers is summarized in Figure 4. While earlier works have attributed the ability of categorical perception to the innateness and uniqueness of human speech module, the capacity to perceive speech contrasts has now been found in several different species (Kuhl 1986), including evidence that chinchillas and other mammals perceive consonant boundaries in a categorial fashion. These findings also provide critical evidence for the current issue in question: if non-human animals can perceive and discriminate speech contrast, speech production must not be a pre-requisite for the ability of speech perception. This renders a strong version of motor theory of speech perception impossible and invalidates the Necessity Condition.

III. Production and Perception of Voice-onset Time (VOT) in Individuals with Speech Disorders

a. Overview of neural correlates of speech production and perception and the types of aphasia

Traditionally, analysis of adult aphasic patients serves as a beginning point of the correlates between speech
capacities and the brain basis of the components of linguistic processing units. The most studied types of aphasia include the Broca's aphasia, also known as the production aphasia, and the Wernicke's aphasia, also known as the receptive aphasia. The Broca's aphasics, resulting from damages to the Broca's area located near the primary motor cortex in the anterior region of the left hemisphere (for most people), often show severe impairments of language production. Broca's aphasia is traditionally analyzed as a speech-motor deficit, labeled the nonfluent aphasia. Patients often demonstrate difficulties implementing specific articulatory gestures, resulting in non-fluent, effortful speech, which its comprehension is often intact.

Wernicke's aphasia involves injury in the Wernicke's area in the posterior region of the left temporal cortex (posterior region of the Superior Temporal Gyrus, which is responsible for speech perception). The Wernicke's aphasics usually has fluent, non-sense speech (with syntactic and semantic impairment) while comprehension is often impaired. Comparing the speech production of Wernicke's aphasics with those of the Broca's aphasics, one finds that they involve different kinds of production errors: while the former is traditionally associated with phonological errors (e.g., substitution of phonological features), the latter is analyzed having more phonetic errors, a result of impaired motor function to implement articulatory gestures. However, the types of mistakes observed in these two kinds of aphasics are not always easily identifiable as phonological or phonetic, as will be discussed below.

b. Nature of VOT production errors in Broca's aphasia and Wernicke's aphasia

Recent literature have favored a more fine-grained analysis of the two types of aphasics by zooming in on the types of production and perception errors involved in the deficits. The result is a far more complex picture which call into question the traditional, oversimplified framework of associating Broca's aphasia with production deficits and Wernicke's aphasia with perception. For instance, it was discovered that Broca’s aphasics do indeed suffer from comprehension deficits: specifically, these patients display problems in the interpretation of sentences when they are forced to rely entirely on grammatical rather than semantic or pragmatic cues. This led to the hypothesis that Broca's area maybe responsible as the “grammar processor” (Caplan & Hildebrandt, 1988; Zurif & Caramazza, 1976). Similarly, the problems associated with Wernicke’s aphasics were reinterpreted to reflect a selective impairment of semantics (resulting in comprehension breakdown and in word-finding deficits in expressive speech), accompanied by a selective sparing of grammar (evidenced by the patients’ fluent but empty speech). Hence Wernicke’s area could be viewed as a “lexical semantic processor” (Ullman et al., 1997). In sum, as Bates et al (1999) put it, there is a 'crisis' in aphasiology that non of the components break down following damage to selective brain areas in a simple enough manner that we can derive a framework for one-to-one mapping from the components of language functions to the components in the brain.

In the light of this fine-grained analysis of aphasics, Blumstein et al (1980) conducted experiments to determine the nature of Broca's aphasics' VOT production errors vs. Wernicke's aphasics' VOT production errors, among other types of aphasics considered. Traditionally, phonological errors are associated with posterior fluent aphasics (Wernicke's aphasics), where as phonetic errors are observed in anterior non-fluent aphasics (Broca's aphasics). However, it was discovered that anterior non-fluent aphasics also show a lot of phonological errors, but it's often
colored by the phonetic errors to be discovered. Meanwhile it is not clear whether the phonological errors is phonological or phonetic in nature: for instance, a /b/-/>p/ error could be either interpreted as a categorical error (phonological), or a lower-level VOT timing error that results in the misperception of the category but is articulatory in nature. Therefore it's not clear whether the surface 'phonological' error in anterior non-fluent aphasics is the same as the posterior fluent patients.

Figure 5 (adapted from Blumstein et al 1980) shows the types errors defined as phonetic and phonemic in VOT production: the first graph shows the production of a normal control, with VOT values of most of the voiced stops below 20ms and most of the voiceless stops greater than 40ms. Here, the two types of errors can be clearly identified: whereas the Broca's aphasics mostly produced abnormal VOT values (mostly between 20-40 ms) that are in general absent from normal controls (phonetic-articulatory errors—meaning that they do not distinguish between the VOT values of the voiced and voiceless categories), the Wernicke's aphasics for the most part simply produced the values in the wrong category (a phonological substitution/mismatch: intended voiced stops-greater than 40 ms, intended voiceless stops-less than 20ms). Further analysis, however, led to the conclusion that even though Broca's aphasics show a predominantly phonetic error, the nature of their errors are both phonological and phonetic (Blumstein 1980:168). Relating these two types of errors to the current discussion on the motor theory of speech perception, it is not clear they show a corresponding relationship with the data on the perception of the aphasics. This will be
c. Perception of VOT in aphasic patients

Blumstein et al (1977) studied the perception of aphasic patients in a more fine-grained framework of testing. In particular, Blumstein et al were interested in the two levels of VOT perception: one test focuses on the mere discrimination of two stops presented, while the other test focuses on the ability to label the perceived sounds. Right-brain damaged non-aphasic patients are tested as controls. Results of the perception tasks indicated that if a subject could not discriminate the stimuli, he could not reliably label them (i.e., no aphasics performed well on the labeling test but poorly on the discrimination test); however, a subject with a normal discrimination function might nonetheless be unable to label the stimuli reliably. These results were interpreted in relation to two levels of processing: one based upon the function of a set of property detectors, the other making use of these properties for linguistic processing. The authors concluded that it is this latter level which seems to be selectively impaired in aphasia.

Particularly relevant to our current discussion is the relationship between the performance on perception tests and the types of aphasia in question, in order to establish association with the production data. We are particularly interested in finding out, in accordance with the production data, whether the Wernicke's aphasics showed a strong tendency of labeling mismatch between phonological categories, while the Broca's aphasics showed a difficulty of discrimination in the first place. Unfortunately, among the types of aphasia studied (including Broca's, Wernicke's, Conduction, mixed anterior, and anomic) the only consistent pattern of deficit among the groups is found in the Wernicke aphasics. Here, three of the four patients were able to perform the discrimination task but were unable to label the stimuli consistently (which is actually consistent with our prediction from the production data). Importantly, though, 4 out of 5 Broca's aphasics demonstrated normal ability to discriminate VOT despite their severe production deficits, and the perception data did not correlate with their production. Other types of aphasics showed a mixed performance on discrimination and labeling conditions. Further experiments on a larger sample size are necessary to determine the correlation between types of perception errors and aphasic types.

d. Evidence for motor theory of VOT perception

Stronger evidence of a causal relationship between motor speech systems and the ability to perceive speech sounds is potentially available from lesion studies. Specifically, if the motor system is critical for the perception of speech sounds, then damage to the motor speech system should have a substantial impact on the ability to perceive speech. For example, Basso and colleagues (Basso, Casati, & Vignolo, 1977) report that 74% of aphasics were impaired on a phoneme identification task and further that the incidence of impairment was greater for non-fluent aphasics (91% of patients) than fluent aphasics (72% of patients). However, another study of non-fluent Broca’s aphasics reports only a mild deficit if any ( 93% accuracy for the most difficult contrasts) on syllable discrimination (Baker, Blumstein, & Goodglass, 1981). However, it has been criticized that these early studies did not have the benefit of modern neuroimaging methods and therefore could not reliably rule out cases with multiple or bilateral lesions, and their methodology often lack controls for response bias (Hickok et al 2011). It is therefore unclear
whether the reported deficits are due to damage to the motor speech system and/or reflect perceptual discrimination or response bias.

e. Evidence for dissociation between VOT production and perception

Following the proposal of the non-necessary mechanism for the motor theory of speech perception (Hickok and Poeppel 2004), Hickok et al (2011) revisited the evidences for the necessary condition (strong version) of the motor theory of speech perception in a series of VOT perceptual-production experiments, which also rather conclusively constitutes a comprehensive review of the dissociation between speech production and perception in lesion study literature. Part of the effectiveness of this study comes from (1) a careful criterion in choosing the 24 radiologically confirmed subjects with Broca's area damage (instead of diagnosed Broca's aphasia symptoms); (2) excluding patients with more diffuse brain damage; (3) effective correction of response bias induced by using percentage correctness (analysis reveals that Broca's aphasics performed better on the different trials than same trials indicating a bias to respond “no”. Given the finding of bias in previous literature, it is unclear whether the performance correlation with Broca’s area damage resulted from receptive speech deficits or from response bias (Hickok, 2010)).

Similar to Blumstein et al (1981) discussed above, this study used auditory VOT discrimination tasks and found that Broca's aphasics' performance was excellent. To assess the functional claim that motor speech ability is critical for speech perception, the study also assessed the relation between fluency and speech discrimination scores on the assumption that fluency is a measure of the integrity of motor speech system as suggested by the correlation between non-fluent speech and frontal lesions (Kreisler et al., 2000); no effect of fluency was found. In addition, no correlation is found between the performance on the discrimination task and the production task. Nonetheless, the current study did find a mild deficiency in the Broca's aphasics' perception score (d'=4.18) comparing to a non-aphasic control group (d'=4.98). Possible mechanisms of this mild impairment are suggested.

The dissociation between production and perception performance in aphasic patients is also observed in a number of other studies in a growing body of literature. For instance, Hoit-Dalgaard, Kopp and Murry (1983) found deficits in both production and perception among apraxia patients but no correlation between the two (“the relationship is neither congruent or predictable”).

f. Other Evidence for dissociation of production and perception

Hickok et al (2011) identified several categories of additional evidence demonstrating that motor speech production is not necessary for speech perception. Combining these evidences with what I found in literature, these involve successful speech perception in conditions of: (i) patients with acquired damage to the motor speech system, i.e., non-fluent aphasics and/or lesions involving Broca’s area (Baker et al., 1981; Blumstein et al., 1977; Rogalsky et al., 2011), (ii) individuals with developmental anarthria (Bishop et al., 1990; Lenneberg, 1962; Weller, 1993), (iii) individuals with acute and complete functional deactivation of the motor speech system via the Wada procedure (Hickok et al., 2008), (iv) healthy infants who have not yet acquired the ability to control their motor speech system (Eimas, Siqueland, Jusczyk, & Vigorito, 1971), (v) non-human animals who do not have the biological capacity for speech (Kuhl & Miller, 1975; Lotto, Kluender, & Holt, 1997), and (vi) individuals with apraxia disorder (Hoit-
IV. Conclusion

In this paper I have considered evidence from developmental studies, studies of normal individuals, and studies of aphasic patients in association with the motor theory of speech perception of VOT. I have demonstrated that while the current evidence does support some form of linkages between VOT production and perception, a strong version of the motor theory of speech perception (the Necessity Condition) is difficult to maintain. The lesion studies in particular provide strong evidence regarding the dissociation between speech-VOT production and perception. Specifically, my conclusion is consistent with the proposal of Hickok and Poeppel (2004), which postulates that the mapping of sensory representations of speech onto motor representations may not be an automatic consequence of speech perception, and indeed are not necessary for auditory comprehension (Hickok & Poeppel 2000), but can be activated and utilized strategically to assist in task performance. Here I offer several concluding remarks here regarding the implication of this finding and for future research.

First, I have argued in the beginning of this paper that motor theory of speech perception was first proposed in order to address the invariance problem of one-to-one mapping of speech sound to mental representations, and that any rejection of the theory will need to propose an alternative to account for this problem. If we abandon the necessity condition, we must also abandon the idea that when we hear the variability in the formant transition from /d/ to various vowels, we do not need the articulatory gestures as a pre-requisite to identify /d/ (which does not have a fixed acoustic characteristic) and to discriminate /d/ from /g/. In a sense, we can ideally examine a 'mute' child, as we have seen in language acquisition literature, and see how his lack of motor speech production affect his perception. However, since the exact mechanisms and degree of muteness in this child is not clearly known, it is difficult to determine his level of perception as compared to the normal definition of 'speech perception'. Nonetheless, alternate proposals to explain how we overcome the problem of invariance has been proposed, such as the Lexical Access From Feature (LAFF, Stevens, 1980s), Examplar Theories (Goldinger 1996), and the TRACE connectionist model (McClelland and Elman, 1986). The evaluation of these models is not within the scope of this paper.

Second, we need to take into account that the majority of evidence that supports a dissociation between production and perception (and negation of the necessity condition) comes from observations of lesion studies, i.e., can production be impaired while the perception is relatively intact, or vice versa? Here we tap into a fundamental logical question in neuroscience with regard to the hypotheses of how observable neurocognitive processes relate to one another underlyingly. In other words, we can ask, can there be logical gaps between the observed order of language breakdown and the hypotheses derived regarding how these functions are acquired, stored/represented, and actively processed? For instance, in order to account for the functional dissociation between music and language processing observed in lesion studies (e.g., aphasia without amusia, or vice versa) and its contradiction with observations of neuroimaging data in normal individuals, Patel (2008) proposed a hypothesis in which language and music share neural resources in their learning mechanism and in their active on-line processing, but once acquired, nonetheless...
have different brain localizations to store their representation. Therefore, it is conceivable that brain injuries may have
damaged brain regions where one of these representations are stored but spared the other, leaving the observer to
conclude that music and language are completely dissociated separate domains whereas in reality they may as well
have shared a considerable amount of neural resources in their processing and acquisition mechanisms (the resource
sharing framework of music and language syntactic processing). This hypothesis was proposed to resolve the
contradiction of evidence between normal individuals and brain-damaged patients; however, it also illustrates the
possible explanations we may consider to account for observable data in lesion studies. Given the overall complexity,
component-integration nature, and the parallel processing capacity of the brain, this consideration is also relevant in
the current discussion of motor theory of speech perception.
Reference


rules are processed by procedural system. Journal of Cognitive Neuroscience, 9(2), 266-276.


