A Two Mechanism Model of Pure Word Deafness

JOSHUA A. RILEY
GREGORY B. COGAN

Abstract

This paper reviews cases of pure word-deafness (PWD) and examines the light they shed on the nature of the mechanisms underlying speech perception. The auditory deficits in each case are examined, including not only speech, but music and significant environmental sounds as well. Audio-visual integration is also considered given the demonstrated utility of visual cues, including lip-reading, for PWD patients. Single-mechanism accounts prove to be insufficient, and a model is proposed which describes the interaction of two mechanisms: a spectro-temporal integrator and a distinct feature extractor. This model helps explain a typology of symptoms in PWD and can be applied to auditory agnosia as well.

Introduction

Pure word deafness (PWD), rare as it is, has attracted considerable attention from investigators interested in speech perception because of its apparent specificity as an impairment of the understanding of speech sounds. The canonical picture of PWD is a patient without impairment in speaking, writing, or reading but who has great difficulty in hearing spoken speech. Moreover, this canonical patient has otherwise normal hearing – the problem appears to be specific to speech, and not the perception of other complex sounds, which is precisely what has made it a matter of import to the debate over the nature of the speech perception system in humans. In the first section, we discuss why PWD should be so interesting to those working on speech perception. PWD patients often display interesting temporal impairments when it comes to perceiving sounds, and in the second section we discuss the question of whether these temporal impairments by themselves can explain PWD. Of course, as with most neuropsychological disorders, there is in practice extensive overlap between PWD symptoms and other auditory agnosias, notably perception of environmental sounds and of music. We discuss these overlaps and their implications in the third and fourth sections respectively. In the fifth section, we incorporate evidence from beyond neuropsychology that has bearing on whatever model of speech perception we will have to posit to account for the patterns of deficits in PWD. In the sixth section we assess the extent to which PWD is truly a domain specific problem, and how speech-specific the mechanisms involved in an account of PWD must be. We go on in the seventh section to propose a model of the perceptual mechanisms we believe to be implicated in PWD, and in the eighth we offer additional evidence to support the model and motivate our separation of all existing cases of PWD into two distinct types.

 University of Maryland Working Papers in Linguistics 16, ed. A. Omaki, I. Ortega-Santos, J. Sprouse and M. Wagers, pp. 201-221. College Park, MD: UMWPiL.

1 The Importance of PWD

Pre-theoretically, the lesion-induced deficit variously referred to in the literature as verbal auditory agnosia or PWD would seem to be the ideal topic of investigation for determining whether or not there is any specifically linguistic about speech perception. The symptoms typically associated with PWD are often idealized as a selective deficit in the comprehension and recognition of auditorially presented words when compared to non-speech sounds, and with intact reading comprehension and spontaneous writing. If it was the case that this idealized, so-called "pure" version of PWD was typical of the condition or well-supported by numerous case studies, the existence of PWD would offer strong support for such a level of processing or representation. If there was a relative impairment for comprehension of spoken words and only for the comprehension of spoken words, and no other stimulus, auditory or otherwise, there would indisputably be something special about speech. Given that the preservation of spontaneous writing and speech and reading comprehension means that some kind of lexical access must be intact, PWD would therefore be considered to be a problem with some special, speech perceptual processing mechanism.

And yet the argument over the inherent specialness of speech continues (or at least, the specialness of those mechanisms that process speech), in no small part due to the fact that "pure" PWD not only is not the typical instantiation of the condition, but also that "pure" PWD may or may not actually exist. Indeed, the sheer rarity of the "pure" version of PWD is such that most documented cases are, in fact, simply the impairment of speech processing capabilities in the absence of general hearing loss, regardless of whether or not there are other, co-occurring complex sound deficits. We will thus follow most of the literature in our use of the term PWD, and refer to the original conception of PWD as "'pure' PWD". An insistence upon only considering cases of "pure" PWD requires discarding a vast majority of the available literature, and we have avoided doing so in the interests of attempting a more comprehensive review of speech processing deficits that are labeled PWD and providing a model to account for more than merely the handful of apparently "pure" cases.

There are certainly case studies that claim to document cases of "pure" PWD (Denes and Semenza 1975; Gazzaniga et al. 1973; Hayashi and Hayashi 2007; Metz-Lutz and Dahl 1984; Takahashi et al. 1992; Yaqub et al. 1988), but the "pure" case studies are often deficient in terms of the sophistication of the tests used to evaluate the patients described. Yaqub et al. (1988), for example, claimed only that their patient could "appreciate" music and could recognize musical instruments, which would not have necessarily required their patient to exploit an intact mechanism of rapid spectro-temporal integration. It is crucial to note that in some sense, the question of the existence of PWD is orthogonal to the question of whether speech is special, since, for example, it is possible that the speech perception system is modular functionally while being widely distributed anatomically. The non-existence of "pure" PWD would simply remove one strong argument in favor of the specialness of human speech, not provide a counter-argument to it.

Still, if it was the case that PWD patients showed relatively uniform and selective impairment for speech perceptual processing, regardless of the kinds of speech stimuli with which they were presented, it would be reasonable to insist upon the existence of special speech mechanisms. After all, the speech signal varies dramatically in its acoustic properties from one point in time to another, and the correlates in the signal of whatever kind of phonological primitive one chooses to acknowledge are radically different. Fricatives, for instance, are identifiable primarily on the basis of prolonged, aperiodic turbulence, while stops are identifiable on the basis of silence followed by turbulence (Ladefoged 2004). Different types of segment thus present different processing challenges for a mechanism focused purely on the analysis of acoustic features of the signal. If processing of these diverse stimuli subtypes were impaired in a relatively uniform fashion, and lexical access itself was still demonstrably intact, there would be very good grounds to assert the existence of some linguistic level of pre-lexical representation independent of the level of acoustic analysis. It is, presumably, only at a linguistic level, and not an acoustic one, that a fricative and a stop can be treated as the same kind of thing.

But time and time again, selective impairment of the processing of different kinds of speech sounds is found in patients presenting with PWD. A fairly robust finding dating to some of the earliest psycholinguistic work on the condition is that PWD patients, despite being relatively good at distinguishing steady-state vowels from each other, perform poorly when asked to discriminate between CV syllables that differ with respects to their consonants (Godefroy et al. 1995; Kazui et al. 1990; Lambert et al. 1989; Miceli 1982; Praamstra et al. 1991; Yaqub et al. 1988). This kind of impairment might have a fairly straightforward account in terms of linguistic representations – most phonological theories make a distinction between vowels and consonants, so it is not unreasonable to think that there is a difference in representation very early on in speech perception – but this is less reconciled with the fact that PWD patients also seem to perform poorly on distinguishing CV syllables when the difference between them lies in the vowel, rather than the consonant (Praamstra et al. 1991). Most mainstream phonological theories would be reluctant to treat vowels that occur without preceding consonants and vowels that are preceded by consonants as functionally different entities, though some phonological theories might be more obviously receptive to such a distinction; Articulatory Phonological approaches (Brownman and Goldstein 1995) would necessarily acknowledge the difference between a vowel gesture that overlaps initially with a consonant gesture of some kind. Since most phonological theories would have difficulty reconciling this breakdown in performance with their basic units of representation, accounts of the deficit observed in these patients that are based on some kind of pre-lexical phonological representation could not be integrated comfortably with these theories. Attempting to substantially change the basic concepts of phonological theory is, perhaps, a gambit that is less appealling than simply withdrawing support from the idea of a specifically linguistic level of representation prior to lexical access in speech perception. Since phonological theory does not fit well with what would be required of a phonological level of representation given the performance of these patients, an alternative that makes no reference to phonology whatsoever has been proposed.

2 **The Crucial Role of Time in PWD**

If PWD is not to be explained by impairment of or damage to some specifically linguistic level of perceptual processing, what mechanism's break-down might be responsible for the symptom pattern observed in PWD? Presumably, the problem must lie in some manner of non-linguistic, acoustic analysis mechanism, since the problem seems to be confined exclusively to the auditory modality. The failure cannot be in low-level audition as such, as audiological testing does not typically reveal profound deafness and complex environmental noises can often still be successfully identified (Buchman et al. 1986; Caramazza, Berndt, and Basili 1983; Griffiths et al. 1999; Metz-Luts and Dahl 1984; Praamstra et al. 1991; Saffran et al. 1976; Takahashi et al. 1992; Tanaka et al. 1987). Similarly, it cannot be the case that there is a total failure at the level of complex sound analysis, since PWD patients have been reported with intact discrimination of auditory stimuli differing in fairly specific acoustic features, such as amplitude and frequency (Wang et al. 2000). Ideally, one particular component of complex sound analysis would explain the symptom pattern of PWD patients, and that component for many researchers has been time, or, at least, temporal integration (Albert and Bear 1974; Auerbach et al. 1982; Miceli 1982).

The initial findings suggesting that impairment responsible for PWD was a temporal stimulus processing impairment were those in Albert and Bear (1974) demonstrating a click fusion effect in PWD patients. Clicks were played for patients and normal controls, who were asked to indicate the number of clicks played in any given trial. While for normal controls, clicks occurring 5-7 msec apart could be clearly distinguished as separate clicks, PWD patients perceived two clicks as being one if they were separated by less than 15 msec. It has since become uncontroversial to assert that there is some problem with temporal resolution in PWD patients, regardless of whether this particular problem fully explains the observed symptoms of the condition. While it is not our claim that meaningful speech transitions usually or often occur within a temporal window of 15 ms, the impairment on this task may indicate more serious failures of temporal integration that affect larger temporal windows with more demonstrable relevance to speech processing.

The temporal account of PWD also explains the frequent finding that simply slowing down the rate of speech helps greatly in a patient's comprehension (Hemphill and Stengel 1940; Klein and Harper 1956; Takahashi et al. 1992). If rapid spectro-temporal integration is what is impaired in PWD, then reducing the difficulty of that task by allowing the patient to perform the required integrations over a longer timescale would necessarily be helpful. This finding, however, is not universal across cases of PWD; several case studies have reported that slowing speech did not significantly facilitate comprehension (Buchtel et al. 1989; Coslett et al. 1984). In such cases, the PWD is generally either particularly severe or accompanied by other deficits (such as the paraphasia noted in Coslett et al. (1984) or the inability to discriminate even environmental sounds in Buchtel et al. (1989)). Slowing down speech may not greatly reduce the difficulty of the integration task, and it may simply be that particularly severe damage to the mechanisms underlying PWD means that slowing down speech isn't sufficiently helpful to achieve success. Alternatively, the co-occurrence of other, non-speech sound-processing deficits suggests that the PWD mechanism is crucially dependent on the continued functioning of other mechanisms, and when those other mechanisms are impaired, it is unable to perform the integration even over a slower time scale.

Perhaps more importantly from the perspective of what is known about patients presenting with PWD, simply saying that the pre-lexical level of linguistic representation is damaged in PWD is not sufficient to account for the peculiar temporal impairments generally displayed. It is not logically necessary that someone who is unable to distinguish /ba/ and /ka/, for example, should also fuse clicks occurring less than 15 msec apart. Any account of PWD should provide some explanation of these peculiar temporal deficits, which means that the critical role of spectro-temporal integration mechanisms must be taken seriously.

3 **Impairments in Environmental Sounds and Speech**

The existence of true deficits in non-speech complex sound identification without impairment for speech would be a very serious challenge to this interpretation of PWD evidence, that speech is simply subserved by highly overlearned algorithms and routines responsible for processing rapid spectral changes at the level of complex sound analysis. To be such a challenge, of course, it would have to be the case that the crucial factor in identifying the sounds on which a given patient was impaired were rapid spectral changes that had to be tracked in time in a relatively narrow frequency band (the *sine qua non* for speech), rather than variations in amplitude, pitch, or frequency changes over a very wide range (true of perhaps most environmental sounds). Only if the environmental sounds on which relatively unimpaired performance is reported are sounds for which rapid spectral changes are crucial for successful identification and discrimination will this potential challenge have been realized. If none of the existing studies that purport to have documented impairment to the recognition of non-speech sounds without similar impairment for speech sounds have used stimuli with these properties, then this particular serious challenge to the idea of speech perception being mediated entirely by a very special kind of complex sound analysis will not have been realized.

On careful examination, the literature does seem to offer some apparent double-disassociations, but they are rare. Far more commonly reported are cases where there is some kind of impairment for environmental sounds, there is an impairment in speech sounds (Barrett 1910; Buchtel et al. 1989; Denes and Semenza 1975; Gazzaniga et al. 1973; Hemphill and Stengel 1940; Kazui et al. 1990; Lambert et al. 1989; Miceli 1982; Nagafuchi and Suzuki, 1976; Praamstra et al. 1991; Reinhold 1950). The converse expected pattern also holds, as there are multiple cases of impaired speech perception in the absence of serious deficit in environmental sound identification (Caramazza et al. 1983; Coslett et al. 1984; Mills 1891; Peretz et al. 1994; Takahashi 1992; Tanaka et al. 1987).

There are a handful of reported cases of patients with an impaired ability to recognize environmental sounds but relatively preserved speech comprehension (Godefroy et al. 1995; Habib et al. 1995; Lambert et al. 1989; Spreen et al. 1965). "Apparent" is used advisedly here; the deficits documented have often not been tested in a particularly satisfactory manner. In Lambert et al. (1989), for example, the patient was tested on her speech sound discrimination ability by being allowed to point at the correct syllable on a multiple-choice matrix, whereas she was required to name the environmental sounds. Godefroy et al. (1995) is similar in reporting a patient impaired on naming environmental sounds, not pointing to them.

It is unclear how consistent this pattern of specific difficulties with naming of environmental sounds is; in Spreen et al. (1965), the deficit reported was not this specific. The patient's deficit was also highly erratic, however, as he would often make mistakes on identifying environmental sounds he had correctly identified minutes earlier, or would be able to give some indication of the significance of sounds without being able to name them. The inconsistent performance of this particular patient suggests a strong attentional component, probably compounded in this case because of the patient's reported hyper-irritability and violent, unprovoked mood swings. The rarity and idiosyncrasy of the deficit patterns in patients claimed to exhibit an environmental sound deficit in the absence of an impairment in speech perception thus makes detailed analysis of what has gone wrong in such cases extremely difficult.

There are also very often deficits in some aspects of normal speech perception in these patients, despite being unimpaired relative to their other problems. Lambert et al. (1989) and Godefroy et al. (1995) both report that their patients were impaired on repeating words presented to them, though this was far more severe for Lambert et al. (1989). We must be careful not to put too much weight on this particular disassociation, at least if it is used to attempt to argue for the existence of specifically linguistic representations. Combined with the murkiness of what is really wrong in this putative double disassociation, it cannot be considered as strong evidence for any particular hypothesis.

Still, given the incredibly frequent co-morbidity of impairments in speech comprehension and environmental sounds, it is worth examining the pattern of impairments in the recognition of environmental sounds displayed by many PWD patients, which often have interesting similarities. When these environmental sound-impaired PWD patients are presented with sounds in a forced or multiple-choice identification task, their performance often improves dramatically, sometimes to within normal levels (Hemphill and Stengel 1940; Miceli 1982; Wohlfart et al. 1952). The assistance of additional information, in this case, a serious constraint on the range of possible analyses to consider, seems to be crucial in successful identification in the face of damage, suggesting a mechanism whose level of performance has been lowered rather than one whose function has entirely ceased.

This is consistent with the finding that PWD patients generally perform far better when presented with additional visual information in the form of lip-reading (Auerbach et al. 1982; Klein and Harper 1956; Lambert et al. 1989; Lichtheim 1885; Metz-Lutz and Dahl 1984; Reinhold 1950). The facilitatory effect of lip-reading is sometimes very dramatic, as in the case of one patient who was able to lip-read well enough to understand newspaper articles read out loud to him! (Wohlfart et al. 1952) The role of lip reading is particularly interesting given the nature of the pre-lexical deficits most often displayed. Grant (2002) proposes that the most pertinent type of information gained from lip reading is that of place of articulation. Greenberg (2006) claims that place of articulation is also the most robust type of information that can be gained from the analysis of the speech signal.

Taken together, this suggests that the usage of specific visual cues (in this case lip reading) can aid in some part in the restoration of the analysis of the speech signal: specifically, for place of articulation. There is still debate about both the nature and the time course of this integration, but the fact of the integration is unquestionable. Provided there is some form of non-auditory information available, performance will improve, suggesting either multiple, mode-dependent routes to lexical access or a modality-integration stage at some point in the speech perceptual process.

4 Amusia and PWD

Another avenue which may force the conclusion that pre-lexical speech perception is not simply the result of tracking rapid spectral changes is the examination of deficits in the perception of music. Much of musical perception does depend crucially on tracking rapid changes in time (Peretz and Zatorre 2005) of precisely the sort that the complex-sound account of PWD relies upon. And it indeed appears to be the case that at least some musical perceptual abilities have remained intact in patients otherwise presenting with PWD, though the kinds of methodological problems noted above undermine the reliability of such findings. The inverse is also claimed in the neuropsychological literature, with documented cases of patients with speech perceptual abilities relatively intact and impaired musical perception (Albert and Bear 1972; Mills 1891; Saffran et al. 1976).

The existence of musical perceptual abilities in patients with PWD would not necessarily be problematic for the complex-sound account *per se*, because the perception of speech certainly involves finer-grained tracking of spectro-temporal changes than the perception of music (Wolfe 2002), with many languages making contrasts between sounds depending on differences of tens of milliseconds, as in the case of voice-onset time distinctions (VOT) (Ladefoged 2004). Given the superior capabilities putatively needed for speech perception on the complex sound-account, however, the inverse case of speech perception without musical perception would be, if sustained, largely irreconcilable with the view of pre-lexical speech perception being a unitary mechanism with complex sound analysis. Without specifying the functional sub-parts of such a mechanism, there is no kind of general damage that would cause a mechanism to perform adequately on the most difficult tasks it is presented with but break down completely on the simpler ones.

As noted above, however, the lack of sophisticated psychoacoustic testing music perception in such patients hampers the drawing of such strong conclusions, generally because clinical investigators have failed to adequately characterize the abilities that have been preserved in such patients. On the basis of such evidence, music perceptual deficits cannot yet have much light to shed on the existence of a pre-lexical linguistic level of representation in speech perception.

To the extent that preserved musical abilities are noted in PWD, these abilities are often fairly uniform across patients. Some ability to identify or imitate melodies is usually preserved (Buchman et al. 1986; Hemphill and Stengel 1940; Saffran et al. 1976; Takahashi et al. 1992; Wohlfart 1952), and very often individual instruments can be identified (Saffran et al. 1976; Takahashi et al.

1992; Yaqub et al. 1988). Thus, the fact that PWD can co-exist with at least some level of musical capability is not disputed. Amusia of some degree, however, often co-occurs with PWD (Miceli 1982; Peretz et al. 1994; Tanaka et al. 1987), even in the absence of notable impairment in the identification of complex environmental sounds (Buchman et al. 1986; Coslett et al. 1984; Takahashi et al. 1992).

Even in the cases cited above as evidence of PWD in the absence of amusia, careful testing usually reveals some relevant deficits, such as poor performance on discriminating melodies (Takahashi et al. 1992) or being unable to tell music apart from speech despite being able to hum melodies (Wohlfart et al. 1952). It is worth nothing that the only case study that has unambiguously asserted the existence of PWD in the absence of any musical deficits whatsoever (Mills 1891) provides no details of the tests used to establish this fact, which greatly undermines its potential to clarify.

In summary, evidence from the amusia literature can have only slight bearing on the question of mandatory pre-lexical linguistic representation in the speech perceptual system. Amusia and PWD seem very often to be co-morbid, and while amusia may involve some of the same kinds of computations as the successful perception of speech sounds, it is unlikely that they are required to perform on the same time scale or level of accuracy as is necessary to distinguish speech sounds from each other in running speech. Thus, the existence of partial musical perception in PWD patients is of limited interest. The converse case, of amusia without PWD, is at least sporadically asserted, but such cases often lack detailed or sophisticated testing of musical perceptual capabilities, as has been noted in previous reviews (Poeppel 2001). While this latter case would certainly suggest at least some language-specific mechanism in play and thus troublesome for the hypothesis that complex sound analysis happens without reference to any kind of linguistic representation, a non-mandatory top-down influence account would not find this state of affairs particularly problematic.

5 Converging Evidence from Beyond Neuropsychology

If it was the case that a specifically linguistic, pre-lexical level of representation played a mandatory role in speech perception in developmentally normal human beings, that level of representation would presumably have as its primitives some manner of specifically linguistic unit. This unit could be the classical phoneme, or the syllable, the distinctive feature, or any proposed purely linguistic, abstract unit. A representational level that did not use some kind of specifically linguistic primitive unit would not in any meaningful sense be a linguistic level of representation. That much is true definitionally.

Furthermore, if this mechanism dealt in such units and this was the level at which these particular units were abstracted from the speech signal, a change in these units should result in some change in the activity of this mechanism. It should therefore be engaged in the task of telling its primitives apart (e.g. distinguishing /ba/ from /bi/) and building or modifying the representations it currently entertains (e.g. putting /ba/ and /tAn/ together in some kind of order). A change in these units should thus lead to more activation of some area selectively activated for perception of stimuli as speech when a change in the acoustic stimuli leads to a change in the perception of one of these units. These are precisely the circumstances in which this level of representation is supposed to be doing work, i.e. when it receives information suggesting that it must change or manipulate the representations and units that it deals with. What should not activate this mechanism is an acoustic signal that does not contain reflexes of whatever linguistic unit it is designed to extract. That sort of activation would suggest that the area underpins at least some mechanisms that are not linguistically-specific, which would be more in line with an account that posited mechanisms that were not linguistically-specific.

There is some evidence to suggest that this mechanism responds even in the absence of the acoustic reflexes of linguistic units. While activation specific to speech versus non-speech conditions in an fMRI study of subjects listening to sine-wave speech was observed in the left superior temporal sulcus, this area, implicated by numerous studies as crucial to speech perception (Hickok and Poeppel 2004**)**, did not show any difference in activation between trials in which a sine-wave stimulus, perceived as speech, was merely repeated and trials in which the stimulus was actually changed to produce the percept of a different phoneme or syllable (Dhaene-Lambertz et al. 2005). A spectro-temporal integration account of the representation of linguistic units in speech perception would predict precisely this; there being no qualitative or systematic quantitative differences in signal complexity between /da da/ and /da ba/, for example, a reliable difference in activation between being presented with a specific linguistic unit and then a second, different linguistic unit would not be expected.

The area that was sensitive to the phonemic/syllabic/distinctive feature difference was the left supramarginal gyrus, commonly activated in syllable discrimination tasks, but also widely argued to subserve higher-order decision or memory processes which may well occur after lexical access or as part of a sensory-motor interface system (Hickok and Poeppel 2004). This would suggest that any linguistic process that depended on the units normally consigned to the intermediate, pre-lexical linguistic level of representation in the speech perceptual stream are properly units only in top-down, post-perceptual mechanisms. Such mechanisms might strongly influence the level of complex analysis such that a representation appropriate for successful lexical access without being a separate, intermediate mechanism as such.

Proposals attempting to define the precise form of these mechanisms have been put forward, supported by EEG evidence based primarily on mistmatch negativity (MMN) responses (Korpilhati et al. 2001; Pulvermüller et al. 2001; for review, see Näätänen 2001). These studies support the idea of language-specific MMN responses to speech stimuli, which strongly suggests that whatever is generating these MMN responses is not simply a complex sound analyzer common to all developmentally normal humans, but something specific to speakers of a given language, and thus probably linguistic in nature. Explicit ideas of what shape these specifically linguistic representations might take can be found in the psychoacoustic literature (e. g. Moore 2003); as yet, however, a model reconciling these proposals has not emerged.

6 Auditory Stimulus Representations Without Domain Specificity?

What might a model of speech perception without mandatory, linguistically-specific pre-lexical levels of representation look like? Given how quickly brainstem auditory evoked potentials recover after encountering a given stimulus, it cannot be the case that anything prior to the complex sound analytical level is responsible for durable representations of stimuli (Picton et al., 1981). Furthermore, the discrepancy between the known time course of the extraction of particular acoustic features of a given signal and the reality of a unified percept means that something must draw together information about all of these features into a unified representation, the features perhaps being united by their co-occurrence within a particular window of time. Psychoacoustic research has demonstrated the existence of such windows through the apparent irrelevance of temporal linear order within such a window for the purposes of establishing a percept or even for signal detection (Carlyon 1987).

The independent existence of precisely this level of representation is supported by a large body of work suggesting that one of the most robust ERPs, the mismatch negativity response, is sensitive to relationships between stimulus events (that is, stimuli occurring in particular patterns in time) rather than stimuli themselves, and conjunctions of acoustic features (for a review, see Näätänen and Winkler, 1999).

Such a representation, with features of an acoustic signal integrated in a particular temporal envelope, could be all that lies between complex sound analysis itself and lexical access. Certainly, it is the case that impairment in integrating so many features into a unified, sensory percept is capable of accounting for cases of PWD as described above and may prove to be useful in understanding all types of auditory agnosia. Indeed, the task of successfully integrating all relevant features may be particularly hard in the case of speech simply because it is normally produced by the human voice. There is neuroimaging evidence that the areas of STS normally implicated in speech versus non-speech paradigms may be sensitive to any sound created by a human voice, whether speech or not, and even when differences in spectral structure are controlled for (Belin et al. 2000).

It may be just these non-spectral cues to the presence of a human voice that activate whatever mechanism mediates the top-down influence of phonological representations in the stage of integration. Asserting this would require successfully identifying the cues to human voices independent of speech and a demonstration that they are processed and thus enter into the emerging integrated representation earlier than information about spectral change, but it is a potentially fruitful avenue for further inquiry.

7 The Model Proposed

Recognizing the need to reconcile the crucial role of spectro-temporal integration in PWD with the fact of lexical access and the existence of phonological processes, our model of the pre-lexical levels of speech perception essentially consists of two mechanisms (see Diagram 1 below). The first of these mechanisms is the temporal stimulus event integrator, *pace* Näätänen and Winkler (1999). Input from simple acoustic feature detectors responsive to derived signal properties like fundamental frequency or amplitude is stored online during a window of time equal in length to at least the time course of the slowest of these feature detectors.

The second of these mechanisms is the distinctive feature extractor. We here use distinctive feature in the traditional phonological sense of distinctive feature, as these are the most widely accepted, sub-phonemic, specifically linguistic unit (Blumstein 1987; Jakobson, Fant, and Halle 1952). The need for some specifically linguistic and sub-phonemic unit is discussed below.

The operation of this second mechanism is crucial to the creation of an actual stimulus event representation, which is the representation thought to correspond to auditory MMN ERPs, as described above. While integration into a stimulus event representation will proceed by itself if no candidates are forwarded by the distinctive feature extractor, if such candidates are supplied, the end product of integration becomes somewhat more uniform, with the result that the stimulus event representation will be far more noisy and likely to lead to an erroneous percept. The integrator compares feature candidates to the acoustic feature information it has received and determines whether it is appropriate. If it is rejected, the extractor attempts to supply another candidate; if it is accepted, then the feature is used to complete the integration into a stimulus event representation.

This final, edited and normalized stimulus event representation is what is then passed forward to the lexicon. In this model, the frequent selective impairment observed for place of articulation rather than manner or voicing in PWD patients (Auerbach et al. 1982; Kazui et al. 1990; Miceli et al. 1982; Yaqub et al. 1988) can be explained by the fact that place of articulation is usually reflected in the speech signal by precisely the sorts of higher-order signal properties (like formant transitions) that can only emerge as a result of the second stage, rather than the first stage of integration. In contrast, manner of articulation is usually indicated by simpler acoustic properties (presence of periodicity, turbulence, sporadic silence) and voicing contrasts simply require knowing at what point periodicity was introduced into a signal. We believe that in PWD patients who do show a selective impairment for place of articulation, the integrator fails to correctly determine whether the feature suggested by the extractor is appropriate, and thus cannot consistently reject erroneous candidates or accept correct candidates. This is demonstrated by the high rate of co-occurrence of significant speech deficits associated with trouble distinguishing place of articulation in phonemic tasks when identification of sounds associated with the first stage of processing is preserved (Auerbach et al. 1982, Miceli et al. 1978; Saffran et al. 1976; Yaqub et al. 1988).

Linguistic units defined on a level larger than a distinctive feature cannot account for differential impairment between place of articulation on the one hand and manner of articulation on the other. This impairment is far below the level of the syllable, and any segment or phonemic representation will have to recognize smaller sub-classes distinguished by manner or place, thus making use of a *de facto* feature representation anyway.

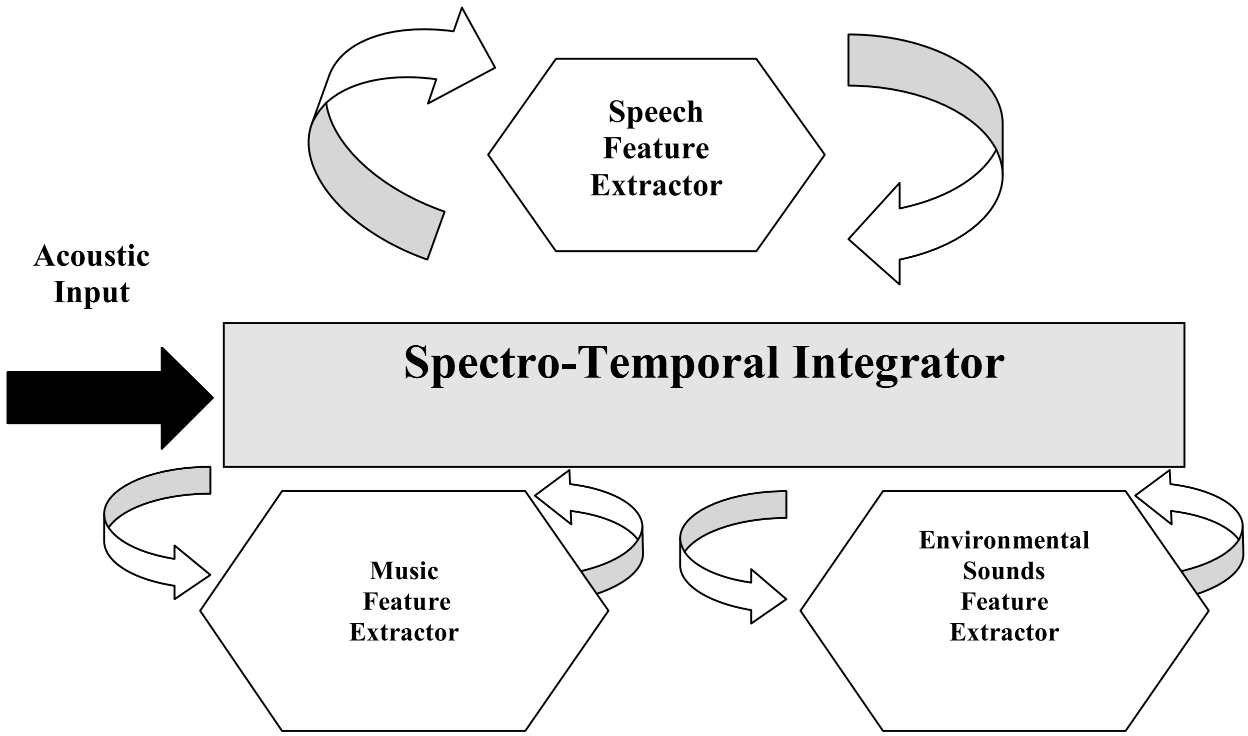

Diagram 1. A Two Mechanism Model of Pure Word Deafness. Input from simple acoustic feature detectors is sent to both the spectro-temporal integrator and feature extractors of various modalities. On the basis of this input, the feature extractors suggest appropriate candidates to the spectro-temporal integrator for completing the integration process. The integrator can reject the proposed candidate as unsuitable, or accept and use the proposed candidate to complete the integration process. The extractor that proposes a successful candidate passes on this fact to post-perceptual, category-determination processes. Damage to the integrator alone leads to the first sub-type of PWD, while damage to the extractors leads to the second sub-type of PWD, discussed below.

8 PWD is not a unitary disorder

Since our model includes two distinctive mechanisms, we must motivate them in the literature. In the more normal case of PWD, the PWD probably characteristic of the majority of patients in the literature, damage has been done to the spectro-temporal integrator rather than the distinctive-feature extractor. These patients generally have impairment in the correct identification of environmental sounds, are amusic to some degree, and have great difficulty comprehending speech. The other, much rarer type of PWD involves an impairment of the distinctive feature extractor itself. These patients present with serious deficits in comprehending spoken words, but are not impaired in their recognition of environmental sounds and are not significantly amusic.

If this second, rarer type of PWD is accepted as damage to the feature extractor rather than the spectro-temporal integrator itself, an account of the rare but real case of patients with deficits in the comprehension of environmental sounds but not speech (Eustache et al. 1990; Godefroy et al. 1995; Lambert et al. 1987), and patients who display amusia without impaired speech perception or environmental sound deficits (Griffiths et al. 1997). These disassociations seem to make it impossible to explain the observed symptoms in the PWD literature by simply invoking some metric of hardness. If PWD's specificity was simply a function of the complexity of the functions performed by one, domain-general mechanism, then for whichever category of sounds is judged to be easiest to process by whatever metric of hardness one might care to propose, there should be no selective impairment for that category and none of the others. Simply put, no general mechanism whose level of performance has simply been decreased should break down catastrophically on the easiest task it has to perform but perform adequately on much harder tasks.

If, however, environmental sounds and music have the same sort of feature extractors, built up on the basis of experience, that exist for speech, the integrator-extractor model accounts neatly for these disassociations. Although these cases are somewhat rare, it should be possible to have damage specifically to these mechanisms, which would produce an impairment not co-morbid with any of the others. This damage might well cause the extractors to generate inappropriate candidates that are forwarded for the integration of the auditory event stimulus representation. If inappropriate candidates are being forwarded to the integrator, the reported inability of some of these second PWD type patients to distinguish music or environmental sounds from speech (Lambert et al. 1989). It is also at least suggestive that it has been reported that some patients with the proposed first PWD type, where amusia is also present, could tell speech and music apart (Auerbach et al. 1982; Buchman et al. 1986), despite these patients' total inability to identify melodies or instruments. Patients with the first PWD type would have an impaired integrator that was failing to correctly process the candidates forwarded to it by the extractors, but because those extractors were successfully forwarding candidates, the patient would be able to identify the category of sound. For patients with the second PWD type, because their extractors are malfunctioning, they would forward candidates regardless of whether or not they had received the right initial cues, so the patient would not be able to easily identify the category of the sound.

This extractor-specific impairment may in practice be distinguished from a simple problem with successfully using candidate features in the integrator by observing whether or not a given PWD patient displays better comprehension of spoken words in the context of sentences or is better able to process isolated words. In the first case, patients will display better comprehension when words are presented in a sentential context (Klein and Harper 1956; Lambert et al. 1989; Roberts and Sandercock 1987; Shivashankar et al. 2001). The syntactic and semantic information available in the sentential context, though severely impoverished from what a normal speaker would receive, can act to push the feature extractor towards suggesting certain candidates based on what the speaker expects to hear next in a sentence. This additional contextual information assists the extractor in suggesting appropriate candidates and in suppressing inappropriate ones, which leads to better comprehension overall.

Those patients, however, for whom sentential context fails to assist (Demes and Semenza 1975) or exacerbates their comprehension difficulties (Praamstra et al. 1991; Yaqub et al. 1988), have PWD of the first type, with damage to the integrator. When the integrator can no longer appropriately make use of candidate features, it does not matter whether that the extractor can forward more and more specific candidate features. As a result, sentential information will become relatively useless, or the integrator will try harder to make use of the more insistently suggested feature candidates, and thus degrade the stimulus representation even further.

It is worth noting that our model refrains from making specific anatomical predictions at present, as the clinical picture normally presented of PWD as a syndrome correlated strongly with lesions to the superior posterior temporal lobe or the connections between those lobes is incomplete. There are two reported patients who presented with the symptoms of classic PWD but whose CT scans revealed that the only damage to their brains were tumors in III ventricle, a subcortical structure whose critical role was confirmed when radiotherapy caused the tumor to decrease in one of these patients and the PWD symptoms to promptly disappear. (Shivashankar et al. 2001). Even more baffling than a PWD case caused entirely by subcortical damage is that reported in Lambert et al. 1989, in which a patient initially presented with an impairment of speech perception but whose MRI scans revealed no obvious damage whatsoever! Anatomical predictions are obviously desiderata of any serious model of a neurolinguistic process, but they are beyond the scope of the present paper.

9 Conclusion

Upon a review of the existent PWD literature, we conclude that there are at least two distinct sub-types of PWD, distinguished by the co-morbidity of word deafness and other auditory agnosias and whether additional information, be it sentential or visual, improves comprehension. Furthermore, these two sub-types of PWD cannot be accounted for by any account based solely on an impairment of spectro-temporal integration. Contrary to the claims of some, time is not a sufficient answer to account for the symptom patterns of pure word deafness.

We propose a two-mechanism model of the stage of speech processing where impairment is found in PWD. Patients with the first sub-type of PWD, who are not assisted by additional information and whose condition is co-morbid with other auditory agnosias, have suffered damage to the first of these mechanisms, the spectro-temporal integrator. This component is not speech specific and therefore damage to this mechanism will have consequences for the perception of all complex sounds, such as certain environmental sounds and music. Patients with the second sub-type of PWD, who are assisted by additional information and do not display significant co-morbidity with other auditory agnosias, have suffered damage to specific distinctive feature extractors. Each one of these sub-types of PWD is the result of damage to a specific mechanism in our proposed model, with the first subtype being the result of damage to the integrator, and the second subtype being the result of damage to an extractor. Damage to both mechanisms is of course possible, which would make separation into different subtypes difficult or impossible, but enough clear cases of damage to only one of these mechanisms exist to justify the separation in principle. Our model finds strong support in work done on audiovisual integration, and is open to the possibility of distinctive feature extractors for certain highly learned non-speech sounds.

References

Albert, M. L., Sparks, R., von Stockert, T., & Sax, D. (1972). A Case Study of Auditory Agnosia: Linguistic and Non-Linguistic Processing. *Cortex*, 8, 427-443.

Auerbach, S., Allard, T., Naeser, M., Alexander, M., Albert, M. (1982) Pure Word Deafness: Analysis of a Case with Bilateral Lesions and a Defect at the Prephonemic Level. *Brain,* 105, 271-300.

Barrett, A. (1910). Case of Pure Word-deafness with Autopsy. *The Journal of Nervous and Mental Disease*, 37, 73-92.

Bauer, R.M. & Zawacki, T. (2000). Auditory Agnosia and Amusia. In M.J.Farahand and T.E.Feinberg (Eds.), *Patient-Based Approaches to Cognitive Neuroscience Issues in Clinical and Cognitive Neuropsychology* (97 -106). Cambridge, MA: MIT Press.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., (2000). Voice Selective Areas in Human Auditory Cortex. *Nature* 403, 309–312.

Blumstein, S. (1987). Speech Perception and Modularity: Evidence from Aphasia. In E. Keller & M. Gopnik (eds.) *Motor and Sensory Processes of Language* 257-289. Hillsdale, NJ: Lawrence Erlbaum Associates.

Buchman, A. S., Garron, D. C., Trost-Cardamone, J. E, Wichter, M. D., & Schwartz, M. (1986). Word Deafness: One Hundred years Later. *Journal of Neurology, Neurosurgery, and Psychiatry.* 48, 488-499.

Buchtel, H.A. & Stewart, J.D. (1989). Auditory Agnosia: Apperceptive or Associative Disorder? *Brain and Language*, *37*, 12–25.

Caramazza, A., Berndt, R. S., & Basili, A. G. (1983). The Selective Impairment of Phonological Processing: A Case Study. *Brain and Language,* 18, 128-174.

Carlyon, R. P. (1987). A Release from Masking by Continuous, Random, Notched Noise. Journal of the Acoustical Society of America, 81, 418–426.

Coslett, H.B., Brashear, H.R., & Heilman, K.M. (1984). Pure Word Deafness after Bilateral Primary Auditory Cortex Infarcts. *Neurology*, *34*, 347–352.

Denes, G., & Semenza, C. (1975). Auditory Modality-Specific Anomia: Evidence from a Case of Pure Word Deafness. *Cortex,* 11, 401-411.

Dhaene-Lambertz G, Pallier C, Serniclaes W, Sprenger-Charolles L, Jobert A, Dehaene S (2005). Neural Correlates of Switching from Auditory to
Speech Perception. *Neuroimage*, 24,21–33.

Engelien, A., Silbersweig, D., Stern, E., Huber, W., Döring, W., Frith, C., Frackowiak, R.S.J. (1995). The Functional Anatomy of Recovery from Auditory Agnosia: A PET Study of Sound Categorization in a Neurological Patient and Normal Controls. *Brain*, 118, 1395-1409.

Fujii, T., Fukatsu, R., Watabe, S., Ohnuma, A., Terumara, K., Kimura, I., Saso, S., & Kogure, K. (1990). Auditory Sound Agnosia without Aphasia Following a Right Temporal Lesion. *Cortex,* 26, 263-268.

Gazzaniga, M. S., Velletri Glass, A., Sarno, M. T. & Posner, J. B. (1973). Pure word Deafness and Hemispheric Dynamics: A Case History. *Cortex,* 9, 136-143.

Godefroy, O., Leys, D., Furby,A., De Reuck, J., Daems, C., Ronde-pierre, P., Debachy, B., Deleume, J.F., & Desaulty, A. (1995). Psychoacoustical Deficits Related to Bilateral Subcortical Hem-orrhages. A Case with Apperceptive Auditory Agnosia. *Cortex*, *31*, 149–159.

Grant, K. (2002). Measures of Auditory-Visual Integration for Speech Understanding: A Theoretical Perspective. *Journal of the Acoustical Society of America,* 112, 30-33.

Griffiths, T. D., et al. (1997). Spatial and Temporal Auditory Processing Deficits Following Right Hemisphere Infarction: A Psychophysical Study. *Brain,* 120, 785-794.

Greenberg, S. (2006). A Multi-Tier Framework for Understanding Language. In S. Greenberg and W. A. Ainsworth (eds.) *Listening to Speech: An Auditory Perspective.* 411-434. Mahwah, NJ: Lawrence Erlbaum Associates.

Habib, M., et al. (1995). Mutism and Auditory Agnosia due to Bilateral Insular Damage - Role of the Insula in Human Communication. *Neuropsychologia,* 33, 327-339.

Hawkins, S. (2003). Roles and Representations of Systematic Fine Phonetic Detail in Speech Understanding. *Journal of phonetics*, 31, 373–405.

Hayashi, K. & Hayaski, R. (2007). Pure Word Deafness due to Left Subcortical Lesion: Neurophysiological Studies of two Patients. *Clinical Neurophysiology*, 118, 863-868.

Hickok, G. & Poeppel, D. (2000). Towards a Functional Anatomy of Speech Perception. *Trends in Cognitive Sciences* 4, 131-138.

Hickok, G., & Poeppel, D. (2004). Dorsal and Ventral Streams: A Framework for Understanding Aspects of the Functional Anatomy of Language. *Cognition* 92, 67-99.

Jakobson, R. C., Fant, G. M., & Halle, M. (1952). *Preliminaries to Speech Analysis.* Cambridge, MA: MIT Press.

Joanisse, M.F. & Gati, J.S. (2003). Overlapping Neural Regions for Processing Rapid Temporal Cues in Speech and Nonspeech Signals. *Neuroimage*, *19*, 64–79.

Jerger, J., Weikers, N. J., Sharbrough, F. W., & Jerger S. (1969). Bilateral Lesions of the Temporal Lobe: A Case Study. *Acta Oto-Laryngologica,* supplement 258, 1-51.

Kazui, S., Naritomi, H., Sawada, T., Inoue, N., Okuda, J. (1990). Subcortical Auditory Agnosia. *Brain and Language,* 38, 476-487.

Klein, R., Harper, J. (1956). The Problem of Agnosia in Light of a Case of Pure Word Deafness. *Journal of Mental Science,* 102, 112-120.

Korpilahti P., Krause C.M., Holopainen I., Lang A.H. (2001) Early and Late Mismatch Negativity Elicited by Words and Speech-Like Stimuli in Children. *Brain and Language*, 76, 332-339.

Ladefoged, P. (2004) *Vowels and Consonants, Second Edition.* Oxford: Blackwell.

Lambert, J., Eustache, F., Lechevalier, B., Rossa, Y., & Viader, F., (1989). Auditory Agnosia with Relative Sparing of Speech Perception. *Cortex,* 25, 71-82.

Miceli, G. (1982). The Processing of Speech Sounds in a Patient with Cortical Auditory Disorder. *Neuropsychologia,* 20, 5-20.

Miceli, G., Caltagirone, C., Gianotti, G., & Payer-Rigo, P. (1978). Discrimination of Voicing Versus Place Contrast in Aphasia. *Brain and Language,* 6, 47-51.

Mills, C. (1891). On the Localisation of the Auditory Centre. *Brain*, 14, 465-472.
Moore, B. (2003) Temporal Integration and Context Effects in Hearing. *Journal of Phonetics*, 31, *563-574.*
Näätänen R, Winkler I (1999). The Concept of Auditory Stimulus Representation in Cognitive Neuroscience. *Psychological Bull*etin, 125, 826-859.
Nagafuchi, M., & Suzuki, J. (1973). Auditory Agnosia Due to Incision of the Splenium corporis callosi. *Acta Oto-Laryngologica,* 76, 109-113.
Näätänen R (2001). The Perception of Speech Sounds by the Human Brain as Reflected by the Mismatch Negativity (MMN) and its Magnetic Equivalent (MMNm). *Psychophysiology*, 38, 1-21.
Peretz, A., Kolinsky, R., Tramo, M., Labreque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional Dissociations Following Bilateral Lesions of Auditory Cortex. *Brain.* 117, 1283-1301.
Praamstra, P., Hagoort, P., Maasen, B. & Crul, T. (1991). Word Deafness and Auditory Cortical Function: A Case History and Hypothesis. *Brain,* 114, 1197-1225.
Picton, T. W., Stapells, D. R., & Campbell, K. B. (1981). Auditory Evoked Potentials from the Human Cochlea and Brainstem. *Journal of Otolaryngology*, 10, 1-41.
Poeppel, D. (2001). Pure Word Deafness and the Bilateral Processing of the Speech Code. *Cognitive Science*, 21, 679-693.
Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Ilmoniemi, R. J., & Näätänen, R. (2001). Memory Traces for Words as Revealed by the Mismatch Negativity (MMN). *Neuroimage,* 14, 607-616.
Roberts, M. and Sandercock, P. (1987). Pure Word Deafness and Unilateral Right Temporo-parietal Lesion: a Case Report. *Journal of Neurology, Neurosurgery and Psychiatry*, 50, 1708-9.
Saffran, E. M., Marin, O. S. M., & Yeni-Komshian, G. H. (1976). An Analysis of Speech Perception in Word Deafness. *Brain and Language,* 3, 209-228.
Shivashankar, N., Shashikala, H.R., Nagaraja, D., Jayakumar, P.N., Ratnavalli, E. (2001). Pure Word Deafness in Two Patients with Subcortical Lesions *Clinical Neurology and Neurosurgery*, 103, 201-205.
Spreen, O., Benton, A., Fincham, R. (1965). Auditory Agnosia Without Aphasia. *Archives of Neurology,* 13, 84-91.
Takahashi, N., Kawumura, M., Shinotou, H., Hirayama, K., Kaga, K., & Shindo, M. (1992). Pure Word Deafness Due to Left Hemisphere Damage. *Cortex,* 28, 295-303.
Tanaka, Y., Yamadori, A., & Mori, E. (1987). Pure Word Deafness Following Bilateral Lesions: A Psychophysical Analysis. *Brain*, *110*, 381– 403.
Ulrich, G. (1978). Interhemispheric Functional Relationships in Auditory Agnosia: an Analysis of the Preconditions and a Conceptual Model. *Brain and Language,* 5, 286-300.

Wang, E., Peach, R.K., Xu, Y., Schneck, M., & Manry, C., II (2000). Perception of Dynamic Acoustic Patterns by an Individual with Unilateral Verbal Auditory Agnosia. *Brain and Language*, *73*, 442–455.

Wohlfart, G., Lindgren, A., Jernelius, B. (1952). Clinical Picture and Morbid Anatomy in Case of "Pure Word Deafness." *Journal of Nervous and Mental Disease*, 116, 818-827.

Wolfe, J. (2002). Speech and Music, Acoustics and Coding, and What Might Music Be 'For'. In Stevens, C., Burnham, D., McPherson, G., Schubert, E., Renwick, J. (eds.) *Proceedings of the 7th International Conference of Music Perception and Cognition, Sydney*. 10-13 Adelaide: Causal Productions.

Yaqub, B.A., Gascon, G.G., Al-Nosha, M., & Whitaker, H. (1988). Pure Word Deafness (Acquired Verbal Auditory Agnosia) in an Arabic Speaking Patient. *Brain,* 111, 457-466.