Psychology of Language and Communication 2012, Vol. 16, No. 1

VERSITAOPEN

DOI: 10.2478/v10057-012-0001-0

ELŻBIETA ŁUKASIEWICZ
Kazimierz Wielki University, Bydgoszcz

PERCEPTION, PROCESSING AND STORAGE OF SUBPHONEMIC AND EXTRALINGUISTIC FEATURES IN SPOKEN WORD RECOGNITION – AN ARGUMENT FROM LANGUAGE VARIATION AND CHANGE

Recent research on speech perception and word recognition has shown that fine-grained subphonemic as well as speaker- and episode-specific characteristics of a speech signal are integrally connected with segmental (phonemic) information; they are all most probably processed in a non-distinct manner, and stored in the lexical memory. This view contrasts with the traditional approach holding that we operate on abstract phonemic representations extracted from a particular acoustic signal, without the need to process and store the multitude of its individual features. In the paper, I want to show that this turn towards the "particulars" of a speech event was in fact quite predictable, and the so-called traditional view would most probably have never been formulated if studies on language variation and language change-in-progress had been taken into account when constructing models of speech perception. In part one, I discuss briefly the traditional view ("abstract representations only"), its theoretical background, and outline some problems, internal to the speech perception theory, that the traditional view encounters. Part two will demonstrate that what we know about the implementation of sound changes has long made it possible to answer, once and for all, the question of integrated processing and storage of extralinguistic, phonemic and subphonemic characteristics of the speech signal.

Key words: speech perception, subphonemic features, extralinguistic features, language variation, sound change

Introduction

One of the fundamental problems of any theory of speech perception and word recognition[1] concerns the primary units of speech processing; are these multidimensional sets of fine-grained phonetic features, or their abstract representations (phonemes), or larger units, like syllables or even words, or, as the proponents of motor

Address for correspondence: Elżbieta Łukasiewicz, Kazimierz Wielki University, Institute of Modern Languages and Applied Linguistics, Grabowa 2, 85-601 Bydgoszcz, Poland. E-mail: el.lukasiewicz@interia.pl

[1] It is beyond the scope of the present paper to discuss the various models of the mental lexicon; therefore, we will use the terms 'speech perception' and 'spoken word recognition' without thereby implying any of the different 'bottom-up' or 'bottom-up plus top-down' models of lexical access.

theories claim, abstract gestural commands? Views on what we actually hear vary. Most contemporary models of spoken word recognition assume that before the level of lexical access, there is a pre-lexical stage in which a phonemic representation of the content of the speech signal is constructed – and these representations (whose nature is a controversial issue in itself) are used in word recognition.[2] In what follows, we will not discuss the problem of the existence of such pre-lexical segmental (phonemic) representations; we will assume that on the interface between auditory and lexical processing there is room for segmental representations. Our attention will be directed to another, closely connected problem – to the question of whether and to what extent subphonemic and extralinguistic information encompassed in the speech signal is perceived and processed by the listener to recover the content of the spoken message and, next, how much of that information is stored in the listener's memory. In other words, the question is whether we process, encode and store the whole range of phonetic cues included in a particular acoustic signal, as the proponents of the episodic approach claim; or, as is claimed by the traditional approach, we operate on abstract phonemic representations extracted from a particular acoustic signal, without the need to process and store the multitude of its individual features.

We might sum up the traditional approach in the following way: the speech perception process is concerned with *what* was said, not *how* it was said. Therefore, also extralinguistic, speaker-specific information is not part of the speech perception process in the proper sense of the term. At present, the dominant view is somewhat different, namely, that fine-grained phonetic and extralinguistic properties of the speech signal are integrally related components of the same acoustic signal and are processed in a non-distinct manner. Although there is not too much research showing exactly *how* and at which stage linguistic and extralinguistic information interplays in spoken language processing, it becomes widely recognized that we need theoretical accounts that will integrate "surface" and linguistic (both phonemic and subphonemic) features of speech.

When discussing the problem of primary units and how much of what we perceive is stored in speech representations in memory, authors like Pisoni and Levi (2007), Nygaard (2005) and others usually contrast the traditional view on speech perception units with more recent findings and point to several problems the traditional approach cannot cope with. These problems concern the invariance and coarticulation effects, as well as the influence of speaker- and episode-specific features on spoken word recognition and word recall. Thus, we might characterize those problems as "internal" to the theory of speech perception.

In the present paper I want to point to a quite different type of evidence – from diachronic linguistics and studies of language change in progress – that can actually give an answer to the fundamental question of whether in spoken word recognition we process and store fine-grained subphonemic and extralinguistic speaker-specific

[2] Cf. TRACE Model, McClelland and Elman (1986); Distributed Cohort Model, Gaskell and Marslen-Wilson (1997); PARSYN, Luce et al. (2000); and others.

information encoded in the original speech signal, or discard it in the process of phonological and lexical analysis as soon as it is no longer needed. Moreover, the evidence that can answer the above question once and for all has been with us for long, and therefore it is only surprising that the so-called traditional view of speech perception should ever have been formulated. We can partly explain this by the relative isolation of diachronic linguistics and sociolinguistics from speech perception studies and the reliance of the latter on abstract phonemic descriptions imposed by the synchronic view of language understood as *langue*, without much interest in *parole*.

In what follows, I will briefly discuss the traditional view ("abstract representations only") and its theoretical background. Next, I will outline some problems, internal to speech perception theory, that the traditional view encounters. In part two, I will show that what we know about the implementation of sound changes has long made it possible to answer the question of integrated processing and storage of extralinguistic, phonemic and subphonemic characteristics of the speech signal.

The traditional view on speech perception units, its theoretical background and limitations

The traditional view has claimed that in the process of speech perception a speech signal is decoded by the listener and stored in the memory by converting the continuous, multi-dimensional, information-rich acoustic signal into a linear sequence of discrete abstract symbols (phonemes) – in the way speech is represented in written form by using broad phonemic transcription, which is, in turn, based on the system of alphabetic writing. Those abstract symbols – invariant and context-free – carry only linguistically significant, contrastive information (the economy principle). As for the innumerable other features present in the acoustic signal (linguistic, subphonemic features and extralinguistic speaker- and episode-specific features), their role in speech decoding is exclusively supplementary (if any), and they are discarded as redundant noise as soon as the right abstract symbol is retrieved. They do not enter the lexical memory store, nor do they have any influence on the set of those abstract idealized representations. Thus, what is variable in the acoustic signal, particularly context- and speaker-dependent, gets "normalized" in the speech perception process (Cf. Joos 1948, Studdert-Kennedy 1976, Halle 1985). Although no one has ever questioned the view that we do perceive and process extralinguistic features such as the tone and quality of voice, accent and dialectal characteristics etc. (sometimes labeled collectively as "surface characteristics" in contrast to "speech content"), that information, according to the traditional approach, is processed separately and has distinct mental representations; it is not part of the speech perception process.

Thus, in the traditional view, the basic units of speech perception do not contain any redundant, episodic information (abundantly present in any speech signal), but only discrete and idealized abstractions. This view was deeply embedded in some assumptions of the structuralist and generative theories of language. In those theo-

ries, generally speaking, the attention was focused on discovering regular patterns in a language system understood as a structure of relations holding between the elements, whereas particular and variable features of those elements were largely ignored. In fact, it was a frequent charge against structuralism that it focused on the orderliness of relational patterns and the objects investigated were *a priori* expected to fit their slot in the pattern, which, in turn, resulted in a greater interest in the relations holding between objects than in the objects themselves.[3] This charge could equally well refer to the "normalization" of the acoustic signal in the traditional approach to speech perception.

Let us briefly outline the reasons why such an abstract, idealized and homogeneous language system became the object of linguists' interest even though it was widely recognized that language used in reality is far from being homogeneous, orderly and invariable. Disregard for the need to describe language in its social and psychological context as a heterogeneous, multidimensional system may have been a result of limitations in methods and tools of investigation in the first half of the 20th century, impeding the work of even those who were interested in the social and psychological aspects of language.[4] However, the main reason why the heterogeneous language system did not become the lawful object of interest for linguistic theory seems to have been de Saussure's *langue-parole* dichotomy and the impact it had on the further development of linguistics. De Saussure defines *langue*, the true object of linguistic studies, as "the social side of speech, outside the individual who can never create nor modify it by himself; it exists only by virtue of a sort of contract signed by the members of a community" (1959, p. 14), whereas speaking (*parole*) is an individual act and reveals individual differences among speakers. Since *langue* is social in nature, i.e. it is common knowledge possessed by every member of a community and it is free of what is particular, accidental and characteristic of individual speakers, it can be studied by linguists even by analyzing their own speech.[5] The final

[3] See Edmund Leach's critical remarks to Lévi-Strauss' *Les structures élémentaires de la parenté*, Leach E. *Lévi-Strauss* (the Polish edition, 1998, p. 122).

[4] The use of anecdotal data, intuitive explanations, and 'thought-experiments' instead of actual data long kept empirically oriented linguistic writing in a kind of informal niche in the otherwise more and more formally oriented 20th-century linguistics. It must be remembered that many of the extralinguistic explanations put forward at the turn of the 19th and 20th centuries, such as those attributing the raising or lowering of vowels to the effects of climate, would strike us today as too bizarre to discuss at all. Consequently, some linguists even underlined the necessity to confine interpretations of language phenomena, for example language change, to purely internal, linguistic factors, and rejected extralinguistic explanations pertaining to dialect geography, psychology, cultural anthropology, sociology etc.

[5] Labov (1972, p. 267) calls it the Saussurean Paradox that *langue*, the social fact, is of homogeneous nature and can be investigated by introspection, whereas what is individual in *langue* calls for sociological research. We can add that as far as the method of introspection in linguistic studies is concerned, de Saussure's standpoint is not quite clear; he writes elsewhere: "If we could embrace the sum of word-images stored in the minds of all individuals, we could identify the social bond that constitutes language. It is a storehouse filled by the members of a given community through their active use of speaking, a grammatical system that has a potential existence in each brain, or, more specifically, in the brains of a group of individuals. For language is not complete in any speaker: it exists perfectly only within a collectivity." (de Saussure, 1959, p. 13-14)

sentence in de Saussure's *Course* says that "... the true and unique object of linguistics is language studied in and for itself" (1959, p. 232). With the caveat in mind that de Saussure's lectures were first published posthumously and this particular sentence might have been an insertion by the editors[6] to underline the general point, it has usually been regarded as a postulate of the autonomy of language study; linguistics should be independent of other disciplines such as psychology, sociology or history. Thus, it was de Saussure who most efficiently removed the social and psychological aspects of language (i.e. the multitude of features found in the language actually used by a speech community and individual speakers) from the mainstream of linguistic studies. The traditional approach to speech perception, though necessarily occupied with the study of *parole*, adopted much of that anti-*parole* attitude by isolating itself from language variation studies.

The belief that it is feasible to work out a fully fledged linguistic theory only if it is based on a view of language understood as a homogeneous and self-contained abstract system gained much reinforcement from the generative theory. Chomsky's distinction between *competence* (abstract knowledge of language rules possessed by the individual) and *performance* (actual use of those rules in speech) clearly reflected de Saussure's earlier dichotomy between *langue* and *parole*[7], and, analogically to *parole*, performance was from the beginning considered irrelevant to linguistic theory. Chomsky's oft-quoted fragment of *Aspects* clearly states that

> *Linguistic theory is concerned with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance.* (1965, p. 3-4)

That Chomsky later replaced the competence-performance dichotomy with the more or less analogical distinction between *I-language* [internal, individual, intensional] and *E-language* [external] is of no consequence for the point presented here, namely generativists' disregard for any systematic treatment of language's heterogeneity. Also in Chomsky's later theories the requirement of linguistic homogeneity is fundamental. The term *E-language*, vaguely defined, serves to comprise all things connected with language use such as performance, utterance, languages understood as social entities – everything that does not deserve to be paid too much attention by the linguist, in contrast to *I-language* which is an orderly, abstract language system internalized in the mind of the ideal speaker-listener.

---

[6] The view first advanced by R. Godel in *Les sources manuscrites du "Cours de linguistique générale" de Ferdinand de Saussure*, Genève, Paris (1957, p. 119, 181) (see the second Polish edition of Cours, 1991, p. 258).

[7] For more information on the analogies between de Saussure's and Chomsky's terminology, see Lyons *Chomsky*, (1998, p. 168-172).

In spite of the merits of structuralism, the *langue-parole* divide was somewhat unfortunate for the theory of language since it furnished it with some insoluble paradoxes – this was aptly pointed out by Weinreich, Labov and Herzog in their seminal article "Empirical foundations for a theory of language change" (1968). But also in psycholinguistics, the traditionally accepted view that speech is perceived, processed and stored in the memory as a linear sequence of abstract, "normalized" symbols – a view relying heavily on the concept of *langue* as a homogeneous, "normalized" system – proved wanting. Soon after the synthetic speech started to be used in research on speech perception, it was demonstrated that it is not possible to reconcile the idea of speech as a linear sequence of discrete symbols with the continuous and multidimensional nature of the acoustic signal. The invariance and coarticulation problems emerged.

To start with, it proved difficult to define a phonetic category as there were no invariant constituents recurring in the acoustic signals heard as instances of the same phoneme – such invariant features of particular phonemes were expected to be found in the acoustic signal, but they were in fact absent (hence the "invariance problem"). The lack of invariance results from the fact that acoustic features of speech sounds depend heavily on their phonetic environment, which varies. Moreover, it turned out that the property of redundancy refers to all acoustic cues, without exception. No acoustic cue, taken in isolation, is really necessary to perceive a particular phonetic category (phoneme), even a cue very characteristic of it. Hence, stops can be perceived as stops without silent periods, we are able to hear fricatives even if the relevant acoustic signal is deprived of friction, or we can perceive vowels without formants, to mention a few examples (Cf. Liberman, Mattingly 1985, p. 11-12).

The second major problem was how to reconcile linearity with coarticulation. The relation between the phonetic units we apparently perceive as phonemes and the stretches of acoustic speech that trigger that perception is far from one-to-one correspondence. The actual articulatory movements do not match the sequence of phonetic categories as those categories appear to us in the phonetic percept (and as they are represented in phonetic transcription). On the one hand, what we call a single phoneme consists of a bundle of articulatory movements which are not simultaneous, and, on the other hand, articulatory movements implied by a sequence of phonemes usually overlap, i.e. acoustic information for a particular phoneme is overlapped with information for another phoneme. So, due to coarticulation, the acoustic signal produced at a given time is influenced by several articulatory gestures simultaneously. The result is that, to take an oft-quoted example (Liberman et al. 1954), the alveolar stop in the syllables [di] and [du] is characterized by completely different second formant transitions. In [di], the F2 transition is high (ca. 2400Hz) and rising to the level of high F2 for [i], but the F2 transition in [du] is low (less than 1200Hz) and falling – to the low level of F2 in the following vowel [u]. Despite the different acoustic nature of the [d]s – different second formant transitions – they are perceived as belonging to the same phonetic category (phoneme), or, in

other words, they sound alike, as the same sound [d]. Why should such different acoustic patterns as rising and falling transitions be categorized under one label?

This has been a problem for any auditory[8] theory of speech perception: if we perceive an acoustic speech signal as a sequence of discrete idealized abstractions (phonemes), then the question arises on what basis that segmentation and categorization are carried out. The absence of invariant features and the lack of linearity due to coarticulation are irreconcilable with the idea that in speech perception we deal with a sequence of abstract representations only.

Another possibility is that we perceive, process and store mental representations which are endowed with much more detailed phonetic information than the traditional view was ready to admit. Goldinger (1998), for example, claims that spoken words are represented in the mental lexicon as sets of very detailed exemplars; those mental representations encompass fine phonetic details as well as surface characteristics (such as those related to the talker's voice quality) of every single exemplar of the word that we happened to encounter. The speed of word recognition depends on the extent to which a word we hear is similar to the exemplars stored in our mental lexicon.[9]

Many other recent studies, for example Pisoni (1993, 1997), Jusczyk (1993), Johnson (1997), Goldinger et al. (1991), Nygaard (2005) provide evidence that we recognize spoken words not as strings of abstract phonemes, but as sets of very fine phonetic features, including speaker-specific details, and so are those sets represented and stored in the lexical memory. "Surface", extralinguistic properties as well as fine-grained phonetic details are integrally related components of the speech signal and are processed in spoken word recognition inseparably from the phonemic information – thus, linguistic processing depends on surface features.

The influence of fine-grained subphonemic cues on lexical access was studied by Andruski et al. (1994) in experiments involving the priming effect under conditions of artificially reduced voice onset times (VOTs) in stops. The priming effect was evident when the priming words contained unreduced, normal VOTs, whereas when VOTs in stops in the relevant words were reduced, the priming effect practically disappeared. This shows that detailed, subsegmental cues influence word processing and lexical access.

In tests on isolated word recognition, Mullennix et al. (1989) showed that, when presented with background noise, isolated words are recognized much more efficiently when subjects listen to data produced by a single speaker than when

[8] In the motor theory, which is not an auditory one, the problem was circumvented. Since the invariants of phonetic categories could not be found in the acoustic signal, the idea was that they are to be found in the underlying motor processes. According to this theory, when we perceive speech we do not perceive sounds but we perceive intended phonetic gestures, and those motoric commands are the locus of invariance – the phonetic invariants are to be found somewhere in the articulatory processes. However, the motor theory has a number of other problems to account for (Cf. Liberman, Mattingly 1985).

[9] For a different contemporary view (words represented as sequences of segments/phonemes consisting of contrastive features only), see Stevens (2005).

they listen to data produced by many different speakers. Analogically, in word repetition tests (word lists presented under intelligible quiet conditions), subjects' results were better when they repeated word lists produced by a single speaker compared to multiple-speaker word lists. This shows that the speech perception system is sensitive to a particular voice quality and must get "retuned" each time the speaker changes.

Being familiar with the speaker's voice matters as well (Cf. Nygaard et al. 1994, Nygaard 2005); we recognize new words more accurately when we hear them produced by voices we know – as compared to stimuli produced by unfamiliar voices. It seems that what we know about a particular speaker's voice, for example his/her vowel tensing, voice onset time, assimilation effects etc., is part of our procedural memory, which facilitates speech processing whenever we hear this particular voice. We do not have to analyze those features anew.

Formant frequencies of particular phonemes are different when produced by different speakers; it is a well-known fact that someone's vowel in *bit* may have the same F1 and F2 values as someone else's vowel in *bet*. This is because the acoustic features of speech sounds depend not only on their phonetic environment, but also on a number of other factors like the speaker's age and sex, the size and shape of the vocal tract, individual voice quality, speaking rate, dialectal variation, pragmatic needs, and many more. However, listeners constructing a phonemic representation and making lexical decisions seem to take such "surface", speaker- and episode-specific properties into account and they change their phonemic classification accordingly. This again suggests that phonemic information and talker information, which are both carried by the same acoustic properties of the signal, are processed inseparably, or in a very close connection.

Summing up, there is ample evidence now that such extralinguistic information like speaker- and episode-specific features is not discarded by the listener as irrelevant at the early perceptual analysis of the acoustic signal. Just the opposite, both linguistic (subphonemic, concerning the features of speech sounds) and extralinguistic (concerning the "particulars" of the speech event) information is perceived, processed and represented in the memory.

Naturally, from the fact that we process and encode phonetic details of particular speech events it does not follow that discrete symbolic representations are no longer needed.[10] That would suggest that language is without duality (double structure) and discreteness – the two features always regarded as fundamental in language architecture. This is not plausible. Segmentation and categorization of the speech signal is performed; it is a reliable process and any speech perception event substantiates this claim.[11] The view espoused by Pisoni and Levi (2007), Goldinger

[10] Such a non-representational view was put forward by Port and Leary in their article "Against formal phonology" (2005).

[11] Cf. evidence from speech errors, such as metathesis or phoneme substitution, speech errors in sequences containing identical phonemes, perception of missing phonemes.

(1998), Nygaard (2005) and many other researchers is rather that both abstract representations and, on the other hand, detailed subphonemic and suprasegmental as well as speaker- and episode-specific "particulars" are processed, represented and stored in the lexical memory (but see Luce and McLennan 2005).

In part two we will turn our attention to research on language change and variation, and will try to show that the view that listeners process and retain fine-grained phonetic and talker-specific information in the memory was an implicit or explicit assumption of many accounts of sound variation and change.

What the traditional view failed to take into account

One of the enduring impacts of de Saussure's legacy is the view that language synchrony is to be kept separate from its diachrony, and that the synchronic description (structural explanation) has priority over the diachronic one (causal explanation) – this view was shared by most of the 20th-century linguistic schools. However, diachronic studies of language change (and variation) can be very illuminating for constructing theoretical accounts of speech perception. The models we construct simply have to be compatible with what we know about actually occurring language variables and language changes – otherwise we create models which are not isomorphic with what they stand for, or worse, which are at odds with reality.

It is an obvious fact that all languages undergo constant change, and the driving force in the development of any language is sound change. In any language or dialect there exist innumerable linguistic variables, some of them very short-lived, others more durable – linguistic variation need not result in a lasting language change, but every language change, and sound change, requires language variation. Those phonetic variables that are involved in a sound change (i.e. do not end up as random variations only) must somehow spread across the lexicon and across the speech community. Although sound changes can be fully understood and described only from a higher-level perspective – that of a speech community over a span of time – it is worth remembering that a sound change takes place in sound systems of individual speakers/listeners. Therefore, the problem of how much acoustic-phonetic information we process and retain in the memory and the problem of implementation of a sound change are closely related.

When we consider a sound change operating within a given period of time, we can discuss it according to three aspects: phonetic (how sound *x* changes into sound *y*), lexical (how the change affects relevant words in someone's vocabulary) and social (how it spreads from speaker to speaker). The first aspect, phonetic, is crucial for our considerations; it draws our attention to the question of whether the change from sound *x* to sound *y* is gradual or abrupt. If it is phonetically gradual and proceeds by extremely small steps, this requires the language user's sensitivity to very fine-grained phonetic information encoded in the acoustic signal and storage of that information in the long-term memory. In fact, that was the view

on sound change already propounded in the 19th century, by the school of the Neogrammarians (Cf. Paul 1880).

Let us briefly outline the major tenets of the Neogrammarian approach. Firstly, a sound change is regular and purely phonetically conditioned; it simultaneously affects all words which include a particular sound in a given phonetic context. Thus, sound changes do not involve non-phonetic factors connected with morphology, syntax or semantics; they operate with necessity, showing no concern for the grammatical consequences. That was the so called "regularity hypothesis" – a view basically supported by later research on sound changes (Cf. Labov 1994). Secondly, a sound change is motivated by greater ease of articulation and tends to affect all speakers of a given speech community simultaneously – a view rebutted by later sociolinguistic research on changes in progress (Cf. Labov 1972). Thirdly, and importantly for the subject of the present paper, the Neogrammarians claimed that sound changes are phonetically gradual, operate by infinitesimal steps, inaudible and unobservable to unaware language users. Sound change is not a single momentary act dictated by convenience. It is only through adding up of a great number of minute displacements motivated by greater ease of articulation that, after a long period, a sound change may result.

Thus, changes proceed by infinitesimal steps[12] impossible for speakers and listeners to notice – but it does not follow that those minute changes in articulation are not *in any way* perceptible to the language user. It is only logical that in order to assure the directionality of a particular sound change which proceeds gradually, the displacements in articulation have to be perceived, processed and stored in the memory. Below we will present how this process was explained by H. Paul in his *Prinzipien der Sprachgeschichte* (1880 1st ed, 1886 2nd ed.) – the Neogrammarians' bible. The account is more of historical value as far as the psychological details are concerned, but the general idea that *all* occurrences of a given sound in a particular phonetic context modify our mental picture of it is quite compatible with Goldinger's (1998) idea (see above) that words are stored in the mental lexicon as sets of exemplars in which all surface characteristics are represented. I do not think that Paul would support the above-discussed traditional view that in speech perception we operate on abstract representations and discard phonetic details as redundant noise. Let us now trace the very mechanism of sound change as it was accounted for by Hermann Paul.

[12] For a different view, see the "diffusionists". According to Wang (1978, p. 238-240), the gradual view of sound change is untenable because too many types of sound changes are incompatible with the idea of imperceptible infinitesimal steps, for example: changes which involve different articulators between which there is no physiological continuum and there is no evidence for phonetically intermediate stages, metatheses, in which sounds x and y are reordered, and flip-flops in which sound x changes into sound y and sound y into x. Also, certain sound changes seem to be operating at a more abstract phonological level and can't be gradual as they involve different simultaneous operations, e.g. a word like *acclimate* in which the pronunciation changed from [əkláımıt] to [ǽklımeıt] – since all three vowels underwent a change in addition to the change in stress, it would be unrealistic to assume that all three vowels shifted gradually, proportionately and imperceptibly.

The two crucial terms in his theory of sound change are *motory sensation* (*Bewegungsgefühl*) and *sound-picture* (*Lautbild*) ([1880] 1978, p. 3). They describe what constitutes the speaker's mental representation of a sound.[13] The motory sensation is formed in the speaker's mind due to certain movements of speech organs involved in the articulation of sounds or groups of sounds. After the direct physical sensation connected with sound production has vanished, a sound-picture (*Lautbild*) is left in the speaker's memory, which is responsible for reproducing similar movements in future and controlling whether the same sound is produced within the same restricted area. The motory sensation (*Bewegungsgefühl*) does not remain unchanged but it is modified by all earlier and current impressions: those identical to and those slightly deviating from the sound-picture (*Lautbild*); they all blend into one. Yet subsequent and, by virtue of this fact, fresher impressions exert a stronger influence on the motory sensation regardless of their frequency. Thus, each change in the motory sensation results in a minute displacement of the limits of possible fluctuations. Paul claims that it is rare for deviations to alternate in their directions so regularly that it would have an overall canceling effect. Normally a deviation to one side dominates (the least effort principle), if only slightly, and soon, with new incoming impressions, a still further change is possible and the resultant minute displacement of the motory sensation ([1880] 1978, p. 8-9).

The possibilities of minute gradual changes in the articulatory movements of speech organs and in the sounds produced thereby might seem unlimited. Yet, there exists a strong barrier to the uncontrolled development of changes in the motory sensation – it is the sound-picture (*Lautbild*). The motory sensation is shaped by the movements and impressions caused by one's own utterances; the sound-picture is also shaped by one's interlocutors. It exercises a controlling power over the motory sensation. Thus, on the one hand, motory sensation is forced to correct itself according to the sound-picture, and, on the other, it cannot fully master the movements of speech organs and gives way to greater convenience. At the same time, a displacement of motory sensation causes a corresponding same-direction change of the sound-picture. In this way the average of fluctuating performances

[13] One might add here that, generally, the 19th-century linguists showed considerable interest in what we might call the psycholinguistic aspects of language and tried to set the regularity of sound change and generally the systematic nature of language in more general psychological principles. According to Paul, any particular unit of language, any class of units and any relation between classes – all have a corresponding image (*Vorstellung*) as their mental representation; these images are associated in groups with multiple interrelations. This constitutes the mental representation of the speaker's linguistic capacity, *psychischer Organismus* as Paul calls it. In order to delineate the history of a language, one must first establish a chain of language states (*Sprachzustände*). The description of such a language state must take into account all the elements of which a language consists and also "it must depict the relation of the elements to each other, their relative strengths, the connections into which they enter, the degree of closeness and strength of these connections." Because the mind of the individual is the locus where all these images and their interconnections are to be found (and examined by self-observation and analysis of one's own *Sprachgefühl*), it is only logical that the language of the individual speaker is the object of linguistic description and, consequently, as Paul writes, "we must distinguish as many languages as there are individuals" ([1880] 1891, p. 35).

may change and yet pass unnoticed because the sound-picture moves together with the motory sensation.

In the light of what has been said, it is clear that sound change cannot be prevented by any conscious effort and it passes unnoticed by the speakers whom it affects in the same way (Paul believed that a speech community is linguistically homogeneous):

> *Of course no such thing as a conscious effort at this result [identity of one's sound production with that of one's interlocutors] exist, but the demand for such agreement remains as something self-intelligible, unconscious.* (Paul [1880] 1978, p. 12)

The Neogrammarians focused on such changes as assimilation and weakening, which involve gradual displacement of the motory sensation; this mechanism of change does not cover sporadic changes like metathesis, epenthesis, haplology or dissimilation. But the problem that only gradient phenomena may change through infinitesimal steps does not seem to bother Paul; non-gradual changes are said to form a relatively small part of the entirety of sound changes ([1880] 1978, p. 21) and are not so much within the focus of Paul's interest. Let us mention here that the so-called "diffusionist" view of the implementation of sound change was markedly different. According to that approach, the Neogrammarian phonetically gradual implementation of sound change is an untenable concept because too many types of sound changes are incompatible with the idea of imperceptible infinitesimal steps (see the footnote above).

For the purposes of the present paper, it is not relevant who was right in the "diffusionists" versus "Neogrammarians" debate, or which type of sound changes dominates: gradual and regular changes or non-gradual and irregular ones (spreading word by word). The occurrence of phonetically gradual sound changes which regularly, by infinitesimal steps, proceed to their completion is a fact, and even if they were in the minority, their sheer occurrence requires an explanation as to how the process takes place. The implementation of gradual, exclusively phonetically conditioned sound changes does require sensitivity on the part of the language user to the fine-grained subphonemic features of the sounds involved in a change and their phonetic context. One has to be sensitive to very detailed acoustic information. Naturally, the implementation of phonetically gradual changes also requires storing that knowledge in the long-term lexical memory of the speaker/listener – there is no other way to account for the gradualness, regularity and directionality of the process. Hence, we not only perceive and process, but also store in our lexical memory a very detailed representation of sounds, with the information about their realization in a particular phonetic context.

That the Neogrammarian view of sound changes outlined above is basically correct (with the exception of the homogeneity postulate) has been supported by a number of recent studies on changes in progress in American dialects of English,

such as the Northern Cities Shift, the Southern Shift, the fronting of /uw/ and /ow/, the fronting and raising of /aw/, and others.[14] All those changes in progress reveal three features characteristic of the Neogrammarian sound change: lexical regularity, phonetic gradualness and phonetic conditioning. On analyzing spontaneous speech (which is considered to show the most regular and consistent sound patterns, without the effects of sporadic self-correcting on the part of informants) it has been observed that in every lexical item, regardless of whether it is common or uncommon, sophisticated or ordinary, the relevant changing sounds move at the same pace and in the same direction. For example, in the process of /ohr/ raising in New York, none of the observed realizations of the sound (when occurring in spontaneous speech) remained at the previous cardinal [o] place, but all tokens moved upward in the direction of [u:e], in more frequent words like *door*, *four*, *for*, *more*, *fork* as well as in the less frequent *born*, *forth*, *fort*, *horns*, *source*.[15] Thus, the change is lexically regular and phonetically gradual.

In those changes in progress, differences in the distribution of particular tokens of a given sound (as revealed in the acoustic F1-F2 diagrams) clearly correspond to the fine interplay of particular phonetic features of the environment, which may favor or disfavor the change respectively. For example, if we analyze (after Labov 1994, p. 182-183, 457-459) the raising of /æh/ in the Northern Cities Shift, we can see that the locations of vowel nuclei in the acoustic diagrams according to their height and peripherality (F1 and F2 values respectively) show the effect of such fine-grained phonetic conditioning. The raising of /æh/ in the vowel system of a female informant from Buffalo, based on a one-hour recording, has shown that the height and peripherality of 26 tokens of /æh/ is clearly governed by their phonetic environments. The strongest influence is exerted by the nasality of the following consonant, hence the /æh/ in *aunts*, *dance*, *hand* is the highest and most peripheral. The second-strongest conditioning factor is the place of articulation. Here, the tokens of /æh/ with following apicals and palatals, *sat*, *mass*, *bad*, *batch*, *old-fashioned* are located higher than /æh/ with following labials and velars, which are grouped in a lower located cluster: *back*, *traps*, *calf*, *track*, *black*. The disfavoring effect of initial liquids is proved by the relatively low position of *last* with initial liquid, and of *traps*, *track*, *glass*, *black* with initial obstruent+liquid clusters, and, notably, by the rather low position of *plant* with initial disfavoring obstruent+liquid cluster in spite of the favoring effect of /n/.[16]

---

[14] For a more detailed view of chain shifts in progress observed in American dialects, see Labov (1994, chapter 6); literature on the subject frequently refers to the results of two research projects conducted at the Linguistic Laboratory of the University of Pennsylvania:
LYS (A Quantitative Study of Sound Change in Progress, 1968-1972. The spectrographic study of patterns of chain shifting in some British and American dialects, reported in Labov, Yaeger, and Steiner 1972)
LCV (Project on Linguistic Change and Variation, 1973-1977. Research on sound changes in progress in Philadelphia, involving the long-term study of 11 neighborhoods and a random survey of telephone users, reported in Labov 1980, 1994).

[15] After Labov (1994, p. 453-456).

[16] Based on the diagrams of phonetic conditioning of /æh/ for Bea White, 54, Buffalo, after Labov (1994, p. 182-183, 457-459).

It follows naturally from such analyses of changes in progress that speakers of the relevant dialect must perceive, process and store in the memory not only the acoustic details of the relevant sound at its present stage of the change (otherwise the change would not be gradual), but they also have to process and retain in the memory the fine details of its phonetic conditioning. This is certainly a thoroughly unconscious process on the speakers' part, but it operates with unmistaken regularity.

Regarding the social aspect of a sound change, i.e. its spread within a speech community, the problem is closely related to our processing and storage of non-linguistic, talker-specific information. The most typical and systematic form of linguistic change, which has the greatest importance for the development of a language system, is the so-called "sound change from below" – i.e. below the level of social awareness.[17] Because the fact of ongoing change is not present in social awareness, in such changes speakers do not control or correct the use of the variable involved in the change – the variable does not show stylistic variation, it has the same value in all contextual styles (formal and informal speech) and, importantly, its spreading in the lexicon is regular; it affects all (phonetically) relevant lexical items. On the other hand, the spreading of such a variable is strictly correlated with the group membership of the speakers: changes from below are stratified by age, gender, social class, ethnic group and neighborhood. Such changes are difficult to trace in their initial stage for native speakers (and also for linguists); speakers do not notice the change for most of its operation period and become aware of it only when it is nearly completed.

If we were to sketch a typical mechanism of a sound change from the social perspective, it might be outlined as follows: It starts within a limited subgroup of the speech community, possibly in response to some social motivations, for example external pressure threatening the identity of this group, as in the well-known case of centralization of /ay/ and /aw/ in Martha's Vineyard, Massachusetts.[18] The linguistic form involved in the change has hitherto functioned as an undefined linguistic variable with irregular distribution. Now it is picked up by all members of the subgroup and the variable spreads regularly to all relevant lexical items. However, the whole process escapes social awareness; it is a change "from below" without stylistic variation. The generation of younger speakers carries the change even further compared to the speech of the originators, and transmits it to the neighboring social groups in the community. Such changes usually originate in the working class, or the lower middle class; the spread of change "from below" to other subgroups depends on the extent to which the others identify themselves with the original group and the values represented by them. The variable involved in the change is now a function of age level, social status, neighborhood etc., i.e. it is correlated with group membership. The change may further expand until it reaches the limits of the speech community. As it proceeds to completion and has

[17] Cf. changes "from above" and "from below" in Labov (1972, p. 123; 1994, p. 78).
[18] Labov (1972, chapter 1), originally published in *Word* 19, p. 273-309 (1963).

managed to affect the whole community, or a sizeable part thereof, the variable becomes a characteristic feature of their speech and the community's members, regardless of social background, show pretty uniform attitudes to the use of that variable, though these attitudes need not emerge "overtly" in the form of open verbal judgments but are often observable solely through unaware reactions.[19] Around this moment the variable acquires social recognition and begins to show stylistic variation: It is regularly present in casual speech but frequently avoided in formal speech. This is so because typically, the originators' subgroup is not the highest social group and the social attitude to the variable is unfavorable since members of the highest social group manage to stigmatize it. The attachment of a "low prestige" tag to the variable prevents speakers from using it in formal contexts, and, eventually, it leads to sporadic and irregular correction towards the high valued speech of the upper class. It is a change "from above." Speakers who adopted and use the stigmatized, low prestige form, strive to eliminate it from their speech (and typically in self-evaluation tests describe their speech as free of that form). When a variable is attributed very low prestige, it may become the subject of overt, disparaging social comment and, finally, disappear from actual speech, surviving solely as a stereotype. If a change originates in a social group of the highest status it is a prestige model for other subgroups, and spreads to other subgroups proportionately to the degree of reciprocal contact (Labov, 1972, p. 178-179).

Since a typical sound change is socially conditioned – it spreads in a speech community gradually according to the socially conditioned patterns outlined above – it is highly likely that speakers process and retain in the memory also non-linguistic speaker-specific information, and processing that information is inseparable from processing the linguistic/phonetic content. Otherwise, that pattern of social conditioning of a sound change would not be observed. Let us remember that the spread of a sound change "from below" to other subgroups depends on the level of their identification with the values of the original group. Therefore, processing and storage of fine-grained phonetic details of the sound involved in a change-in-progress, as well as its phonetic conditioning, must be integrally connected with the processing of non-linguistic talker-specific acoustic information in order for the sound change in question to be phonetically gradual, phonetically conditioned and socially patterned.

Conclusions

The traditional view of the speech perception process has held that the initial acoustic signal, which is continuous, multi-dimensional and information-rich, is converted into a linear sequence of discrete, abstract representations (phonemes)

[19] For example through Lambert's "matched guise" techniques and other methods, see Labov (1972, p. 207-215, 248-250).

and those abstract representations – invariant and context-free – carry linguistically significant, contrastive information only. Other features present in the acoustic signal (linguistic, subphonemic features as well as extralinguistic speaker- and episode-specific ones) are discarded as redundant noise as soon as the right abstract symbol is arrived at. This model of speech perception was profoundly influenced by the Saussurean idea of *langue* understood as a homogeneous, "normalized" language system. However, it proved inadequate in the face of problems with invariance and coarticulation. More recent psycholinguistic research on speech perception and word recognition has shown that fine-grained subphonemic as well as speaker- and episode-specific characteristics of a speech signal are integrally connected with segmental information; they are all most probably processed in a non-distinct manner, and stored in the lexical memory.

In fact, such a turn towards the "particulars" of the speech event was quite predictable, and the so-called traditional view would most probably have never been formulated if studies on language variation and change had been taken into account when constructing models of speech perception. The mechanism of a typical sound change outlined above leaves no doubt that in speech perception we are not only very sensitive to phonemic information responsible for the recovery of the message content, but we also process and store in the memory the whole range of fine-grained subphonemic and extralinguistic information – it is not discarded as redundant noise. The fact that sound changes are usually gradual, phonetically conditioned and socially patterned greatly supports Goldinger's model of the mental lexicon (1998), in which word representations are not abstract, idealized, context-free entities but collections of particular exemplars, rich with fine phonetic details, and including "surface", talker-specific characteristics.

References

Andruski, J. E., Blumstein, S. E. & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52, 163-87.

Chen, M. Y. & Wang, W. S-Y. (1975). Sound change: actuation and implementation. *Language*, 51, 255-292.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Galantucci, B., Fowler, C. A. & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13 (3), 361-377.

Gaskell, M. G. & Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Process*, 12, 613-656.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-279.

Goldinger, S. D., Pisoni, D. B. & Logan, D. B. (1991). The nature of talker variability effect on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, 152-162.

Halle, M. (1985). Speculation about the representation of words in memory. In V. Fromkin (Ed.) *Phonetic Linguistics* (pp. 101-114). New York: Academic Press.

Johnson, K. (1997). Speech perception without speaker normalization: an exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing*. (pp. 145-166). San Diego, CA: Academic Press.

Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 363-389). Oxford: Blackwell.

Joos, M. A. (1948). Acoustic phonetics. *Language*, 24, supplement 2, 1-136.

Jusczyk, P. W. (1993). From language general to language specific capacities. The WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 3-28.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1981). Resolving the Neogrammarian controversy. *Language*, 57, 267-308.

Labov, W. (1994). *Principles of Linguistic Change, Volume 1: Internal Factors*. Oxford: Blackwell.

Labov, W. (2001). *Principles of Linguistic Change, Volume 2: Social Factors*. Oxford: Blackwell.

Labov, W., Yaeger, M. & Steiner, R. (1972). [LYS] *A Quantitative Study of Sound Change in Progress*. Philadelphia: U.S. Regional Survey.

Leach, E. (1998). *Lvi-Strauss*. 3rd edn. Translated by Piotr Niklewicz. Warszawa: Prószyński i S-ka.

Liberman, A. M., Delattre, P. & Cooper, F. S. (1954). The role of consonant-vowel transitions in the perception of stop and nasal consonants. *Psychological Monographs,* 68, 1-13.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory revisited. *Cognition*, 21, 1-36.

Luce, P. A., Goldinger, S. D., Auer, E. T. & Vitevitch, M. S. (2000). Phonetic priming, neighborhood, activation, and PARSYN. *Perception and Psychophysics*, 62, 615-625.

Luce, P. A. & McLennan, C. T. (2005). Spoken word recognition: the challenge of variation. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 591-609). Oxford: Blackwell.

Lyons, J. (1998). *Chomsky*. 3rd edn. Translated by Barbara Stanosz. Warszawa: Prószyński i S-ka

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 10, 1-86.

Mullennix, J. W. & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*, 47, 379-390.

Mullennix, J. W., Pisoni, D. B. & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85, 365-378.

Nygaard, L. C. (2005). Perceptual integration of linguistic and nonlinguistic properties of speech. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 390-413). Oxford: Blackwell.

Nygaard, L. C., Sommers, M. S. & Pisoni, D. B. (1992). Effects of speaking rate and talker variability on the representation of spoken words in memory. *Proceedings 1992 International Conference on Spoken language Processing*, (pp. 209-212). Banff, Canada.

Nygaard, L. C., Sommers, M. S. & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42-46.

Paul, H. ([1880] 1891). *Principles of the History of Language.* Translated by H. A. Strong from the 2nd ed. (1886) of *Prinzipien der Sprachgeschichte.* Halle: Max Niemeyer.

Paul, H. ([1880] 1978). On sound change. In P. Baldi & R. N. Werth (Eds.) *Readings in Historical Phonology: Chapters in the Theory of Sound Change.* (pp. 3-22). University Park Philadelphia: The Pennsylvania State University Press.

Pisoni, D. B. (1993). Long-term memory in speech perception: some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication*, 13, 109-125.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.) *Talker Variability in Speech Processing.* (pp. 9-32). San Diego, CA: Academic Press.

Pisoni, D. B. & Levi, S. V. (2007). Representations and representational specificity in speech perception and spoken word recognition. In M. G. Gaskell (Ed.) , *The Oxford Handbook of Psycholinguistics* (pp. 3-18). Oxford: Oxford University Press.

Port, R. & Leary, A. (2005). Against formal phonology. *Language*, 81, 927-964.

Saussure, F. de (1959). *Course in General Linguistics.* Translated by Wade Baskin. New York: McGraw-Hill.

Saussure, F. de (1991). *Kurs językoznawstwa ogólnego.* 2nd edn. Translated by Krystyna Kasprzyk. Warszawa: Państwowe Wydawnictwo Naukowe.

Stevens, K. N. (2005). Features in speech perception and lexical access. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 125-155). Oxford: Blackwell.

Studdert- Kennedy, M. (1976). Speech perception. In N. J. Lass (Ed.) *Contemporary Issues in Experimental Phonetics* (pp. 243-293). New York: Academic Press.

Twaddell, W. F. (1952). Phonemes and allophones in speech analysis. *Journal of the Acoustical Society of America*, 24, 607-611.

Wang, W. S. Y. (1978). Competing changes as a cause of residue. In P. Baldi & R. N. Werth (Eds.), *Readings in Historical Phonology: Chapters in the Theory of Sound Change*. (pp. 236-259). University Park Philadelphia: The Pennsylvania State University Press.

Weinreich, U., Labov, W. & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann & Y. Malkiel (Eds.), *Directions for Historical Linguistics*. (pp. 95-195). Austin: University of Texas Press.