Title: Phonetics (An Overview)
Author: Natasha Warner
Address: Box 210025, University of Arizona, Tucson, AZ 85721-0025
Email: nwarner@arizona.edu
Phone: 520-626-5591

Abstract:
This entry defines phonetics and explores overlap with various applications and with other fields. It explores what constitutes a phonetic theory and what theories are currently common in the field. Overlap and differences with phonology are addressed in particular. It considers major recent changes in the field, and concludes by asking what comes next for phonetics.

Keywords: phonetics, articulatory, acoustic, speech, perception, laboratory phonology, corpus phonetics

Key points:

- The article attempts to define phonetics.
- It discusses the difference between phonetics and phonology, and their overlap, but does not attempt to identify a dividing line.
- The article discusses major theories of phonetics such as exemplar models of speech perception, Articulatory Phonology, and Motor Theory, as well as theories of learning closely related to phonetics such as SLM and PAM.
- The article points out that many phonetics papers test specific, detailed theoretical questions about theories of a single part of the speech system, rather than supporting or arguing against one over-arching named theory of the whole system, as is more common in phonology.
- The article addresses major recent changes in phonetics, particularly changes in methods.

I. Introduction

Phonetics of spoken languages encompasses the study of how the vocal tract produces speech, what properties that speech has in the sound wave, and how listeners understand the speech they hear. (Phonetics of signed languages studies similar things, but in the visual medium of signs (Tyrone, 2020).) Most speech is produced by creating various constrictions in the vocal tract from the glottis to the lips while pushing air out of the lungs and changing those constrictions over time. (This includes constriction by adduction of the vocal cords to produce voicing and constriction or opening at the velum to create nasalization or its absence, as well as constrictions at various locations in the oral cavity that produce [p, ʃ, j, l, i, ɚ] etc. (Ladefoged & Johnson, 2014).) The air flowing out from the lungs through these various constrictions gains various patterns of very fast changes to air pressure: periodic vibrations (voiced portions), high frequency aperiodic vibrations (frication, aspiration, or burst noise), silent portions (stop closures), and changes among these patterns over time. All of these very fast, tiny changes of air pressure constitute the sound wave, and this is the information that reaches the ear of the listener. These changes that convey the linguistic information can be as fast as approximately 0.25 ms (for [s] frication noise around 4000 Hz). It seems somewhat amazing that humans can use our vocal tracts to produce these very specific changes in air pressure, and also amazing that listeners can interpret them so easily that this is the primary means we use to convey information, unless we use a signed language. Phonetics covers all steps of this process: how speakers produce the string of sounds, the properties of the sound wave they produce, and how listeners interpret the sound wave.

II. Overlap of phonetics with other fields, with a focus on phonology

Phonetics overlaps with or helps to inform many fields. Phonetics in combination with psycholinguistics covers how listeners process sounds, words, and sentences over time to extract information (e.g. how the listener gets from the sound wave to the words and meaning when hearing "coat," "coach," "coating," "coding," or "What's the code?", a set that, in American English, demonstrates phonetic differences between stops, affricates, flaps, voicing, monomorphemic vs. inflected or derived words, and question vs. statement intonation). Such processing has both language-specific and general components (Cutler, 2012, all references cited in this article will be single examples of relevant research rather than a literature review, given the scope of the topic). Phonetics together with speech science and software engineering can develop speech synthesis and automatic speech recognition (ASR) systems. Phonetics in combination with knowledge of pathologies and treatments can help speakers with speech or hearing disorders, through the field of Speech, Language, and Hearing Science (Adler-Bock et al., 2007). Hall-Lew (2024) discusses overlap of phonetics with mental health treatment. Forensic phonetics forms part of criminal justice. Phonetics is useful in Applied Linguistics for teaching pronunciation to foreign language learners. Phonetics in combination with sociolinguistics and linguistic anthropology helps us understand how speakers express their identities through speech (Wong & Hall-Lew, 2014). Phonetics in combination with historical linguistics, overlapping into sociolinguistics, helps us understand language change over time and reconstruct past language varieties (Labov et al., 2013). Phonetics can play a role in language revitalization, depending on the goals of the particular language community. This role can involve understanding how to pronounce phonetic transcriptions in old archival documents (Hinton, 2001) or could potentially involve using waveforms and spectrograms to learn how to pronounce unfamiliar distinctions as fluent speakers or speakers on archival audio recordings pronounce them. (However, one should be cautious about whether such methods are actually helpful for the community's goals.) Perhaps closest to home, phonetics together with phonology provides us with a more sophisticated understanding of language at the sound level, as demonstrated by the Laboratory Phonology (LabPhon) conferences and publications.

Many linguists have addressed the question of what constitutes phonetics vs. phonology. This article will not attempt to provide an answer to this complex question. At a basic level, phonology asks questions about the structure of sounds in language and about how sounds relate to each other, while phonetics asks questions about how those sounds are produced (articulatory phonetics) and perceived, and what acoustic patterns define them (acoustic phonetics). Phonology is more often categorical, while phonetics uses more continuous scales. The question of what syllable types can participate in reduplication in the world's languages is clearly phonological, whereas the question of what distribution the frequencies of noise have in a particular fricative of a language is clearly phonetic. However, the boundaries between these things are usually unclear.

One very positive development over the past few decades is that phonetics and phonology have approached each other, increased their overlap, and then merged to such a degree that they now seem to be almost the same field. Anecdotally, from the author's perspective, in the 1990's these two subfields were still rather separated, with formal phonology concerned almost exclusively with representations, rules, constraints, and how the mechanisms within formalisms were allowed to work (Kenstowicz, 1994). Phoneticians were working on a variety of questions, some of them quite theoretical (e.g. Articulatory Phonology and Motor Theory of speech perception, see below), but kept their work largely separate from formal phonologists' abstract formalisms. The Laboratory Phonology conference (first held in 1987, first proceedings as Kingston & Beckman, 1990, see also Ohala & Jaeger, 1986) began to push for using phonetic methods to answer phonological questions and viewing phonology more broadly. Some formal phonologists began to seek out experimental evidence to support their theoretical claims (Flemming, 2001). Over the course of the early 2000's, this mutual approach of the two subfields accelerated. It is now somewhat difficult to find a talk at a conference that is about purely abstract formal theory of sound structure without experimental data (e.g. proposing a particular OT constraint in a language). It is also relatively rare to find one that is about purely descriptive phonetics of a particular sound in a particular language (e.g. documentation of a language's particular fricative), without asking some theoretical question about language structure. Neither phonetics nor phonology research looks the same as it did before the two fields approached and nearly merged. However, the author estimates informally that the change from formal phonology has been larger than the change from phonetics. In the author's judgement, this merger into laboratory/experimental phonology has been a good thing for the field and for progress in understanding both language and speech.

III. Theories of phonetics

One can also ask what makes theories themselves phonetic vs. phonological. Optimality Theory (OT) and earlier theories such as Autosegmental Phonology are primarily phonological theories, although some authors (Boersma, 2011) have attempted to use OT to model continuous phonetic differences in sounds. At the other extreme, exemplar models of speech perception seem thoroughly phonetic (see below). In between are theories such as Articulatory Phonology (Browman & Goldstein, 1992) and Motor Theory of speech perception (see below). Despite the word "phonology" in the name, AP is probably considered mostly a theory of phonetics, dealing with how speakers adjust the relative timing and size of gestures in the vocal tract. Theories of intonation and prosody are by nature both phonetic and phonological, because modern intonation research is so closely tied to the fundamental frequency (f0) curve, but by nature asks what categorical structure creates that curve (Beckman & Pierrehumbert, 1988). Jun (2022) shows how the ToBI (Tone and Break Indices) system serves both as a theory of what prosodic distinctions are possible and as a descriptive prosodic transcription system. In some ways, the International Phonetic Alphabet itself is a theory of what sound distinctions are possible in languages.

Exemplar models of speech perception (also called episodic models, and sometimes extended to speech production as well) generated and influenced a particularly large amount of phonetic research in the first decade of the 21st century (Johnson, 1997). Specific exemplar models differ, but the overall idea is that when listeners hear speech or any other sound, they store the acoustic measurements of the specific exemplar of sound they hear. The acoustic dimensions stored need to cover all acoustic cues the listener uses to perceive both speech and non-speech sounds, e.g. measures that capture vowel quality, voicing, type of frication noise, voice quality, fundamental frequency, spectral tilt, nasalization, etc. That is, each token of the word "the" or "dog" or perhaps each token of the sound /d/ or /ə/ that the listener has ever heard is stored as many acoustic measurements of one token (exemplar) in the brain, along with acoustic measurements of every honking horn, barking dog, or beeping computer the listener has ever heard. (The model may allow for forgetting over time to account for the large number of past exemplars.) The listener also stores the information about which categories each past token was recognized as belonging to. Linguistic categories could be something like phonemes (this exemplar was recognized as /d/), words (this exemplar was recognized as "dog"), something larger (common phrases), or variable. Exemplars are also categorized as non-linguistic categories such as "elderly speaker" or "German accent" or "female," and this information is also stored for each past exemplar. New incoming sound is measured on the same acoustic dimensions, and the new token activates categories that nearby stored exemplars in the multidimensional space were recognized as. Thus if the new incoming exemplar's acoustic measurements are similar to past exemplars that were recognized as /d/ or "dog" or "male speaker," it will activate those categories, and if those categories are more strongly activated than the categories for /b/ or "dug" or "female speaker," the former categories will be recognized, and the new exemplar will be stored as connected to those categories. Exemplar models are often viewed as building sociolinguistic effects directly into the model, since linguistic and non-linguistic categories are recognized and stored in the same way. One advantage to exemplar models is that they have a fully implemented mathematical model, which allows for very specific testing of whether effects that are claimed to support exemplar models actually do show up in the modeling. However, exemplar models are very powerful and can model a great many effects, so research that claims to support exemplar models should include careful thought about what results could falsify the predictions.

Some older theories continue to underlie research in phonetics, such as Articulatory Phonology and Motor Theory. These two theories are radical for linguistics in proposing that the lexical entries of words do not consist of any sort of phoneme-like unit, but rather of articulatory gestures and their timing. For example, the lexical entry for "span" would include a tongue tip raising gesture overlapped with a labial closing gesture and a glottal opening gesture, which are timed closely to create the initial consonant cluster; a tongue back opening gesture that creates a widened pharyngeal opening throughout the word, a velic opening gesture spanning the vowel and final consonant, and a shorter tongue tip raising gesture toward the end of the velic gesture to make the final [n] (Browman & Goldstein, 1992, p. 158). This representation does not include any phonemes /spæn/.

Some other theories are not specifically theories of phonetics, but have motivated a very large quantity of phonetically based research, such as the Speech Learning Model, Perceptual Assimilation Model, and the recently updated L2LP (van Leussen & Escudero, 2015). These are theories of learning in bilingual speakers, not theories of phonetics per se, but they fit well with phonetic research.

Most phonetics research sets out to answer a particular question, rather than to support or refute a particular named theory. One could think that many phonetics papers are descriptive or light on theory. However, they contain theoretical questions about specific parts of the speech and language system. To name a few disparate examples, Honda and colleagues model tongue motion for vowel articulation (Honda, 1996), Steffman et al (2022) ask how listeners use hierarchical prosodic structure to evaluate temporal cues to specific speech segments, and Shattuck-Hufnagel and colleagues show how acoustic landmarks define important portions of the speech signal for the listener (Yun et al., 2020). It seems to the author that most phonetic papers address specific theoretical questions rather than overarching theories of the whole system.

IV. Recent major changes in phonetics

There have been several major shifts in phonetics within the past few decades, such as the increase in "corpus phonetics." This uses large corpora of sometimes relatively spontaneous speech, various degrees of automated measurement of the speech signal, automated forced alignment to locate the segments to be measured, and linear mixed effects models or Bayesian modeling for statistical testing (Liberman, 2019). At its extreme, this method allows researchers to feed large quantities of rather unstructured audio into a forced aligner followed by Praat scripts to extract measurements (Boersma & Weenink, 2024), without ever examining the waveform or spectrogram, potentially without even listening to the speech. At less of an extreme, these methods can speed up researchers' work on a large set of recordings of wordlist reading, with hand-checking of all automatic measures. These methods have allowed phoneticians to measure far more speakers and tokens than they could have with older "small data" methods. Although automatically measured data will contain some errors, one can remove some proportion of the errors through outlier exclusion. However, there is a danger with automatic measurement that one does not notice things in the speech that would affect the results. For example, if a substantial proportion of the speech contains creaky voice, automatic measurements of fundamental frequency will reflect the proportion creaked rather than raising/lowering of pitch. If outlier exclusion procedures do not lead the researcher to notice that the f0 distribution is bimodal (creak vs. modal voice), one might not notice that the automatic measurements are not reflecting pitch. However, automatic measurement methods have allowed the field to examine large quantities of speech in more natural interactions, helping the field move away from reliance on wordlists in "Now I say X again" frame sentences. These methods have also allowed more investigation of sociophonetic topics that need to be studied with many speakers over many regions or decades to address how sound changes progress through society and the lexicon (Labov et al., 2013, Wong & Hall-Lew, 2014, Sonderegger et al., 2017).

The field has shifted toward more sophisticated statistical methods (e.g. LME, Bayesian modeling, GAMMs), and this has affected all of phonetic research, not just corpus phonetics. Within the past two decades, there has been a nearly complete shift from ANOVAs and occasional pairwise correlations to these methods. The level of statistical expertise among phoneticians has increased, as has the frequency of statistics courses being offered within Linguistics departments and the number of statistics textbooks focusing on language (Sonderegger, 2023). However, it is not yet clear that phoneticians on the average have enough statistical expertise to use these methods effectively to reach reliable conclusions. Issues include how to handle repeated measures factors and random slopes (Barr et al., 2013) and what to do for failure to converge (with Bayesian modeling as one solution, Vasishth et al., 2018). The lack of consensus on statistical methods has also caused problems in journal article reviewing. Still, these methods have proven useful, especially for large corpus phonetics datasets. GAMMs are especially useful for data that forms continuous curves over time or space, such as tongue shapes or fundamental frequency curves (Wieling, 2018).

In the last few decades, phonetics has increased the focus on variability in speech. Phoneticians have long known that speech is highly variable, even within one speaker's productions of the same word. However, recent work on reduced speech, especially in spontaneous conversation, has brought a greater appreciation of just how variable the speech we use for daily life communication is (Johnson, 2004; Ernestus & Warner, 2011). Examples such as "did you eat yet" produced as [dʒiʔjɛt] have long been known, but when we readily find examples like "she wants to be a" pronounced as [ʃɨzɪ] or "Friday night" as [fɹɛ̃ː], we start to realize that the extent of phonetic variability in speech may be greater than we had thought (see https://sites.arizona.edu/nwarner/reduced-speech-examples/ for audio examples). Another topic that has grown in recent years is individual differences, investigating factors such as personality, cognitive ability in various types of processing, weighting of acoustic cues, and the longer-known effects of aging, hearing loss, and second language proficiency (Yu & Zellou, 2019; Schertz et al., 2015).

One new method has changed articulatory phonetics greatly within recent decades: the use of ultrasound for speech research (Mielke et al., 2017). Ultrasound has opened up opportunities to measure what the tongue is doing in real time with less discomfort for the speaker than the older electromagnetography (EMA) method. Ultrasound can be taken to the field, and is sometimes usable with children or endangered language elders. Ultrasound is particularly effective for articulations in the back of the oral cavity, which usually cannot be observed with EMA. While all methods have limitations and difficulties in how to measure the data, ultrasound has greatly increased our understanding of speech articulation (cf. the Ultrafest conference series).

V. Conclusion

What is next for the phonetics of the future? First, it is clear that the field needs to find ways to balance and integrate the results from "big data" automatic measurement and "small data" methods. This would allow us to answer questions both with detailed certainty of what happened in individual tokens and with the more representative data that large corpora can bring. Liberman (2019) projects that ASR and automatic measurement methods may become good enough to do most of the work of the phonetician on ever-larger datasets. Second, we need to reach consensus about statistical methods, at least to the point of agreeing that a wide variety of methods can be used to produce valid conclusions. We also need to continue to increase statistical teaching and competence within the field. A third direction to examine is the ongoing convergence of phonetics and phonology. Beyond these areas, the author will be interested to see what might await phonetic research in upcoming years.

References

Adler-Bock, M., Bernhardt, B. M., Gick, B., & Bacsfalvi, P. (2007). The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech Language Pathology,* 16, 128-139.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Beckman, M. E., & Pierrehumbert, J. (1988). *Japanese Tone Structure*. Cambridge: MIT Press.

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. In A. Benz, & J. Mattausch (Eds.), Bidirectional Optimality Theory (pp. 33-72). (Linguistik Aktuell/Linguistics Today; Vol. 180). John Benjamins Publishing Company. https://doi.org/10.1075/la.180.02boe

Boersma, Paul & Weenink, David (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.14, retrieved 22 July 2024 from http://www.praat.org/

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155-180.

Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*. MIT Press.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, *39*(SI), 253-260.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology*, *18*(1), 7-44.

Hall-Lew, L. (2024). Phonetic cues to depression: A sociolinguistic perspective. *Language and Linguistics Compass*, e12529. https://doi.org/10.1111/lnc3.12529

Hinton, L. (2001). The use of linguistic archives in language revitalization: The Native California Language Restoration Workshop. In *The Green Book of Language Revitalization in Practice* (pp. 419-428). Brill.

Honda, K. (1996). Organization of tongue articulation for vowels. *Journal of Phonetics*, *24*(1), 39-52.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J.W. Mullenix (Eds.), *Talker Variability in Speech Processing*, 145-165. Morgan Kaufmann Publishers.

Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium* (pp. 29-54).

Jun, S.-A. (2022) “The ToBI Transcription System: Conventions, Strengths, and Challenges” (pp. 151-181), in J. Barnes & S. Shattuck-Hufnagel (Eds.) *Prosodic Theory and Practice*. The MIT Press.

Kenstowicz, M. J. (1994). *Phonology in Generative Grammar* (Vol. 7). Cambridge, MA: Blackwell.

Kingston, J. C., & Beckman, M. E. (Eds.). (1990). *Papers in Laboratory Phonology: Volume 1, Between the Grammar and Physics of Speech* (Vol. 1). Cambridge University Press.

Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, 30-65.

Ladefoged, P., & Johnson, K. (2014). *A Course in Phonetics.* Cengage Learning.

Liberman, M. Y. (2019). Corpus phonetics. *Annual Review of Linguistics*, *5*(1), 91-107.

Mielke, J., Carignan, C., & Thomas, E. R. (2017). The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: An ultrasound study. *The Journal of the Acoustical Society of America*, *142*(1), 332-349.

Ohala, J. J., & Jaeger, J. J. (Eds.). (1986). *Experimental Phonology* (p. 239). Orlando, FL: Academic Press.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183-204.

Sonderegger, M. (2023). *Regression Modeling for Linguistic Data*. MIT Press.

Sonderegger, M., Bane, M., & Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language*, 598-640.

Steffman, J., Kim, S., Cho, T., & Jun, S. A. (2022). Prosodic phrasing mediates listeners' perception of temporal cues: Evidence from the Korean Accentual Phrase. *Journal of Phonetics*, *94*, 101156.

Tyrone, M. (2020). Phonetics of sign language. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.

van Leussen, J. W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*, 1000.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147-161.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86-116.

Wong, A. W. M., & Hall-Lew, L. (2014). Regional variability and ethnic identity: Chinese Americans in New York City and San Francisco. *Language & Communication*, *35*, 27-42.

Yu, A. C., & Zellou, G. (2019). Individual differences in language processing: Phonology. *Annual Review of Linguistics*, *5*(1), 131-150.

Yun, S., Choi, J. Y., & Shattuck-Hufnagel, S. (2020). A landmark-cue-based approach to analyzing the acoustic realizations of American English intervocalic flaps. *The Journal of the Acoustical Society of America*, *147*(6), EL471-EL477.