Consonant intelligibility of speech produced by dyadic rule synthesis

Joseph P. Olive and Louis C. W. Pols

Citation: The Journal of the Acoustical Society of America **66**, S23 (1979); doi: 10.1121/1.2017665
View online: https://doi.org/10.1121/1.2017665
View Table of Contents: https://asa.scitation.org/toc/jas/66/S1
Published by the Acoustical Society of America

tionally. In the present study, we have investigated several modifications to the spectral folding method with the goal of reducing or masking the tonal noises at the cost of increased computations. One modification requires randomly perturbing the nonzero samples of the spectrally folded signal, while another involves preflattening of the baseband. A third approach performs HFR in the DFT frequency domain. The paper describes these techniques and presents the results of both informal and formal speech quality evaluation of the new and existing HFR techniques. [Work sponsored by the Department of Defense.]

2:20

K2. Lingua: A language interpreter used for demisyllable speech synthesis. Catherine P. Browman (Bell Laboratories, Murray Hill, NJ 07974 and New York University, New York, NY 10003)

A set of programs, collectively called Lingua, have been developed to apply rules to an input string for a speech synthesis system. Lingua is a general purpose system, in which the user specifies the basic unit size for the synthesis, the set of input characters, any set of features, and any type of rule that can be expressed in linguistic notation. Lingua was used to synthesize speech from the demisyllable inventory established by Lovins [Lovins, Macchi, and Fujimura, J. Acoust. Soc. Am. Suppl. 1 **65**, S130(A) (1979]. The system will be described and the resulting speech demonstrated.

2:35

K3. Application of linguistically-motivated rules of syllabication to automatic speech synthesis. Daniel Kahn (Bell Laboratories, Murray Hill, NJ 07974)

One can derive a set of rules for the syllabication of English words and phrases using traditional linguistic argumentation—for example, simplicity of the resulting grammar, agreement with native speaker intuitions in clear cases, etc. In this paper, I describe a practical application of such rules, the automatic syllabification of phonetic strings as a component of a text-to-speech synthesis scheme whose basic unit of concatenation is the demisyllable [Fujimura, J. Acoust. Soc. Am. Suppl. 1 **59**, S55(A) (1976)]. Some important characteristics of these linguistic rules are the use of a "maximal-initial-cluster" principle, no constraints on final clusters, and the assignment of "ambisyllabicity" in certain cases (an "ambisyllabic" consonant is one which is simultaneously the final element of one syllable and the initial element of the next syllable). Informal listening tests have lent support to the hypothesis that these linguistically motivated rules are appropriate in the practical application of syllable-based speech synthesis.

2:50

K4. Consonant intelligibility of speech produced by dyadic rule synthesis. Joseph P. Olive (Bell Laboratories, Murray Hill, NJ 07974) and Louis C. W. Pols (Institute for Perception TNO, Soesterberg, The Netherlands)

We measured the intelligibility of speech generated by a speech synthesis system which converts text to sound by rules. The synthesis process uses dyadic concatenation and is intended for automatic voice answerback in telephone directory assistance situations. The intelligibility of initial and final consonants in monosyllabic words was measured. For comparison two other conditions were tested: PCM-coded speech (12 bits, 10-kHz rate) and LPC-resynthesized speech (12 parameters).

The overall percentage correct score for initial and final consonants, averaged over the 33 listeners, was 93.0% and 92.8% for PCM-coded speech, 86.5% and 85.4% for LPC-resynthesized speech, and 58.2% and 73.5% for rule-synthesized speech. Each word list was preceded by two practice runs consisting of one example word for every initial or final consonant. After this very short training phase, the subjects showed no more learning during the actual word list. The voicing character of initial voiced fricatives was not well represented in the rule-synthesis system. For most other identification errors it was clear that a more optimal choice of the dyads would cause a great improvement.

3:05

K5. The use of LPC coefficient interpolation to improve the naturalness of sentence formation by word concatenation. Arvin Levine (Institute for Mathematical Studies in the Social Sciences, Ventura Hall, Stanford University, Stanford, CA 94305)

A recent experiment [Sanders *et al.*, J. Acoust. Soc. Am. Suppl. 1 **64**, S1(A) (1978)] showed that an important contributor to the unnaturalness of sentence formation by word concatenation was the use of individually recorded words (isolation forms). Experience at Stanford has shown interpolation and quantization of LPC formant frequencies and bandwidths to be effective as methods of data rate reduction. Another use for interpolation is to improve the perceived naturalness of synthesized speech by smoothing the transitions between independently recorded sounds. Several considerations for successful word-boundary interpolation are: (1) syntactic, semantic and phonetic conditions, (2) the interpolation curve—linear or more complex, (3) the durations and locations of the interpolation region in individual words, and (4) the use of a specially prepared "demisyllabic bridge" to connect words. The relative importance of smoothing transitions involving high-frequency function words (articles, auxiliary verbs, etc.) is compared to interpolating all the junctions in a sentence. [Work supported by NSF grant SED–77–09698.]

3:20

K6. Data compression of Linear Prediction (LP) analyzed speech by interpolation and quantization of pseudo-formant parameters. William Sanders, Carolyn Gramlich, Arvin Levine, Robert Laddaga, and Patrick Suppes (Institute for Mathematical Studies in the Social Sciences, 192 Pine Hall, Stanford University, Stanford, CA 94305)

Pseudoformants and bandwidths, derived from LP coefficients, prove highly susceptible to data compression techniques such as interpolation and quantization because of their low entropy and smooth sequential nature. We have developed a heuristic to label initially unordered pseudoformants in a way that reveals their sequential smoothness, so that linear interpolation can be successfully applied to each pseudoformant separately. A linear least-squares fit is made over several points; the number of points is limited by a bound on the allowed mean-squared fractional difference between the data and the interpolating line, and another bound on the maximum-squared fractional difference. Quantization of the resulting parameters is constrained not to introduce more error than was allowed in the interpolation procedure. All parameters need not be interpolated using the same bound; if the relative importance of the parameters is chosen in an acoustically reasonable way, interpolation can produce very high quality sound at considerably reduced bit rates. [Work supported by NSF.]

3:35

K7. Perceived speech quality of Linear Prediction (LP) analyzed speech compressed by means of interpolation and quantization of pseudoformant parameters. Carolyn Gramlich, William Sanders, Arvin Levine, Robert Laddaga, and Patrick Suppes (Institute for Mathematical Studies in the Social Sciences, 192 Pine Hall, Stanford University, Stanford, CA 94305)

It is possible to produce high quality LP speech with a low bit rate by interpolating pseudoformants and bandwidths derived from the