

# Coaching Teknis Penulisan Proposal Penelitian Tahun 2022

PPM, Telkom University, 2021-09-29



Dr. Suyanto, S.T., M.Sc.

Web: <http://suyanto.staff.telkomuniversity.ac.id>

Email: [suyanto@telkomuniversity.ac.id](mailto:suyanto@telkomuniversity.ac.id)

HP/WA: +62 812 84512345

Fakultas Informatika, Telkom University



- **Teknis Review Proposal dari Aspek Reviewer**
- **Teknis Penulisan Proposal dari Aspek Pengusul**
- **Studi Kasus Proposal Lolos Pendanaan**

- **Penilaian proposal dilakukan oleh dua reviewer (atau tiga jika terpaksa):**
  - Internal Universitas
  - Eksternal Universitas
  - Setiap reviewer dibekali **Akun SCOPUS** oleh DIKTI
- **Penilaian proposal dilakukan pada dua bagian:**
  - Rekam jejak **ketua peneliti** (bukan anggota) (40 poin)
  - Penilaian substansi proposal (60 poin)
  - **Total Skor sekitar 55 poin untuk Lolos Pendanaan**

## 1 Jurnal internasional bereputasi

Tidak memiliki publi  
Memiliki publikasi be  
Memiliki publikasi di  
Memiliki publikasi di  
Memiliki publikasi di

Publikasi jurnal/prosiding yang mendapatkan skor:

- Hanya publikasi ketua peneliti (bukan anggota)
- Penulis utama (first author sekaligus corresponding)
- Penulis pertama (first author)
- Penulis korespondensi (corresponding author)

## 2 Jurnal internasi

Tidak memiliki publi  
terindeks sebagai pe  
Memiliki publikasi ju  
sebagai penulis pert  
Memiliki publikasi ju  
sebagai penulis pert  
Memiliki publikasi ju  
sebagai penulis pert  
Memiliki publikasi ju  
sebagai penulis pert

- **Penulis pendamping (coauthor) tidak mendapatkan poin !!!**
- **Untuk buku ber-ISBN: semua penulis mendapatkan poin**

## 3 Buku ber-ISBN

Tidak memiliki publi  
Memiliki publikasi be  
Memiliki publikasi be  
Memiliki publikasi be  
Memiliki publikasi be  
Memiliki publikasi berupa > 3 buku (skor = 10)

- **Skor Pasti** karena berdasarkan data faktual
- **Tidak ada perbedaan** skor antar reviewer
- **Maksimalkan skor 40 poin**

ternasional  
onal terindeks  
onal terindeks  
onal terindeks  
onal terindeks

This author profile is generated by Scopus [Learn more](#)

# Suyanto, Suyanto

Telkom University, Bandung West Java, Indonesia [Show all author info](#)

56843751100 [Connect to ORCID](#)

- [Edit profile](#)
- [Set alert](#)
- [Save to list](#)
- [Potential author matches](#)
- [Export to SciVal](#)

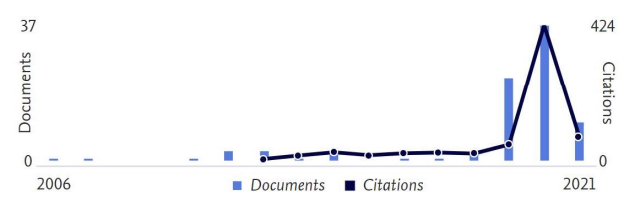
**Metrics overview**

87  
Documents by author

683  
Citations by 279 documents

14  
h-index: [View h-graph](#)

## Document & citation trends



[Analyze author output](#) [Citation overview](#)

## Most contributed Topics 2016–2020

- Cuculidae; Fireflies; Chiroptera  
[14 documents](#)
- Grapheme; Phonemes; Syllabification  
[14 documents](#)
- Speaker Verification; Language Recognition; Utterance  
[3 documents](#)

[View all Topics](#)

87 Documents [Cited by 279 Documents](#) [0 Preprints](#) [102 Co-Authors](#) [Topics](#) [0 Awarded grants](#)

[Export all](#) [Save all to list](#)

Sort by: [Cited by \(highest\)](#)

# Analyze author output

About analyze author tool ?

[Back to author details page](#)

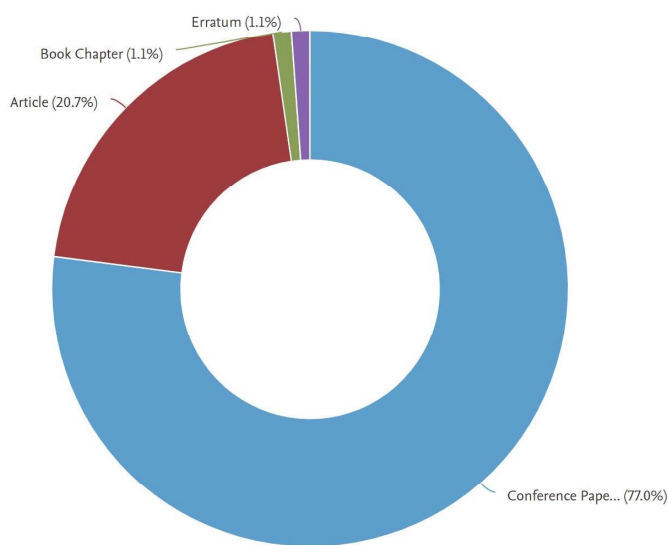
[Export](#) [Print](#) [Email](#)

## Suyanto, Suyanto

Telkom University, Bandung West Java, Indonesia  
Author ID:56843751100

Document type ↓	Documents ↑
Conference Paper	67
Article	18
Book Chapter	1
Erratum	1

### Documents by type



87

by source



by year



by subject



Establishing secure connection...

IF-41-GAB02 (2021....csv)

IF-41-GAB02 (202....html)

Bonferroni\_Mean\_....pdf

Novita\_ChineseCh....pdf

Bonferroni\_Mean\_....pdf

Novita\_ChineseCh....pdf

Show all

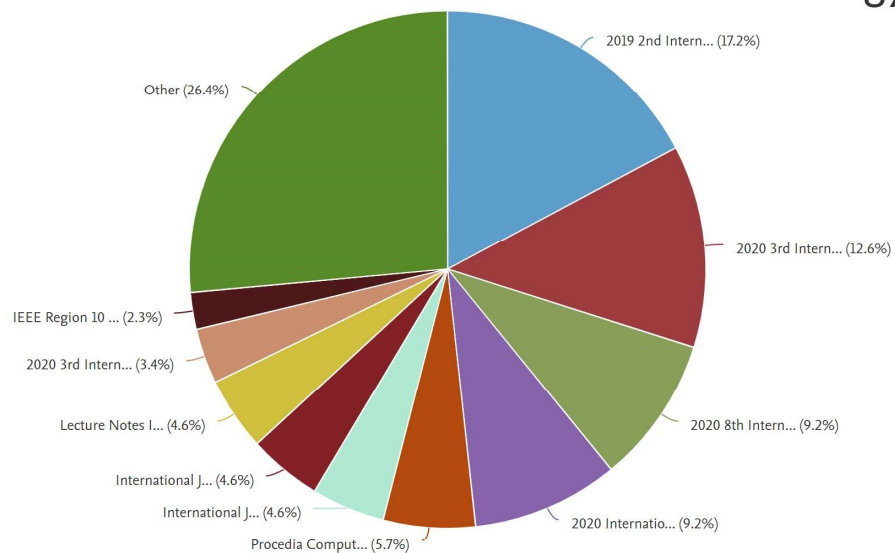
# Suyanto, Suyanto

Telkom University, Bandung West Java, Indonesia  
Author ID:56843751100

Source ↓ Documents ↑

Source	Documents
Proceeding Series	
IEEE Access	1
International Journal Of Computing	1
International Journal On Electrical Engineering And Informatics	1
Journal Of Computational Science	1
Journal Of ICT Research And Applications	1
Journal Of King Saud University Computer And Information Sciences	1
Optik	1

## Documents by source



87

by type



by year



by subject



Click on cards below to see additional data.

No	Komponen Penilaian	Skor Maksimum
1	Relevansi usulan penelitian terhadap bidang unggulan, tema, dan topik Renstra	3
2	Kualitas dan relevansi tujuan, permasalahan, state of the art, metode dan ketepatan waktu	10
3	Keterkaitan dengan penelitian sebelumnya	5
4	Kesesuaian dengan target TKI	5
5	Luaran penelitian	10
6	Luaran penelitian	10
7	Kewajaran biaya	5
8	Kesesuaian target TKI	3
9	Kesesuaian jadwal penelitian	3
10	Kekinian dan sumber primer pengacuan pustaka	5
11	Dukungan kerjasama penelitian	1
<b>Total</b>		<b>60</b>

Komponen penilaian yang mendapatkan skor terbesar:

- Kualitas dan kebaruan riset (10 poin)
- Target luaran (**kelayakan janji**) tahun 1 (10 poin)
- Target luaran (**kelayakan janji**) tahun 2 (10 poin)
- **Skor Tidak Pasti** karena berdasarkan justifikasi reviewer
- **Banyak perbedaan** skor antar reviewer



# Mengapa Tidak Lolos?

- **Rekam jejak ketua peneliti (bukan anggota) kurang kuat (hanya 10 dari maksimal 40 poin)**
  - First/Corresponding author di jurnal internasional bereputasi 0 (0 poin)
  - First/Corresponding author di jurnal internasional bereputasi 1-2 (5 poin)
  - First/Corresponding author di jurnal internasional bereputasi 3-5 (10 poin)
- **Penilaian substansi proposal (hanya 30 dari maksimal 60 poin)**
  - Kualitas dan kebaruan riset (5 poin)
  - Target Luaran (kelayakan janji) Tahun 1 (5 poin)
  - Target Luaran (kelayakan janji) Tahun 2 (5 poin)
  - Komponen lainnya (15 poin)
- **Total Skor < 50 (60/70 bergantung kompetisi) poin sehingga Tidak Lolos Pendanaan !!!**
- **Bagaimana solusi dan strateginya? Tingkatkan rekam jejak dan substansi proposal**

- **Proposal Penelitian Dasar**

- 2019 – 2020: Pengenalan Ucapan Audiovisual Bahasa Indonesia Berbasis Deep Learning
- 2021 – 2022: Deteksi Penyakit Glaukoma Multikelas Menggunakan Dynamic Ensemble Classifier

- **Proposal Tesis Magister**

- 2020 – 2020: Pembacaan Gerak Bibir Berbasis Silabel Bahasa Indonesia Menggunakan Deep Learning

- **Proposal World Class Research (WCR)**

- 2021 – 2023: Model Silabifikasi Fonemis Kata dan Entitas Nama Bahasa Indonesia

- **Proposal Penelitian Terapan**

- 2019 – 2020: Aplikasi Transkripsi Suara dan Klasifikasi Komplain Pada Layanan Pelanggan Indihome Berbasis Deep Learning
- 2021 – 2022: Silabifikasi dan Fonemisasi untuk Model Text-To-Speech Bahasa Indonesia yang Lebih Akurat dan Natural

## 1. JUDUL PENELITIAN

Pengenalan Ucapan Audiovisual Bahasa Indonesia Berbasis Deep Learning

Bidang Fokus RIRN / Bidang Unggulan Perguruan Tinggi	Tema	Topik (jika ada)	Rumpun Bidang Ilmu
Teknologi Informasi dan Komunikasi	Pengembangan sistem berbasis Kecerdasan buatan	Pengembangan aplikasi sistem cerdas	Teknik Informatika

Kategori (Kompetitif Nasional/ Desentralisasi/ Penugasan)	Skema Penelitian	Strata (Dasar/ Terapan/ Pengembangan)	SBK (Dasar, Terapan, Pengembangan)	Target Akhir TKT	Lama Penelitian (Tahun)
Penelitian Kompetitif Nasional	Penelitian Dasar	SBK Riset Dasar	SBK Riset Dasar	3	2

## 2. IDENTITAS PENGUSUL

Nama, Peran	Perguruan Tinggi/ Institusi	Program Studi/ Bagian	Bidang Tugas	ID Sinta	H-Index
SUYANTO Ketua Pengusul	Universitas Telkom	Teknik Informatika		28705	5
KURNIAWAN NUR RAMADHANI S.T, M.T Anggota Pengusul 1	Universitas Telkom	Teknik Informatika		6015401	0
SATRIA MANDALA S.T, M.Sc., Ph.D Anggota Pengusul 2	Universitas Telkom	Teknik Informatika		6026574	4

### 3. MITRA KERJASAMA PENELITIAN (JIKA ADA)

Pelaksanaan penelitian dapat melibatkan mitra kerjasama, yaitu mitra kerjasama dalam melaksanakan penelitian, mitra sebagai calon pengguna hasil penelitian, atau mitra investor

Mitra	Nama Mitra
-------	------------

### 4. LUARAN DAN TARGET CAPAIAN

#### Luaran Wajib

Tahun Luaran	Jenis Luaran	Status target capaian ( <i>accepted, published, terdaftar atau granted, atau status lainnya</i> )	Keterangan ( <i>url dan nama jurnal, penerbit, url paten, keterangan sejenis lainnya</i> )
1	Buku Hasil Penelitian	sudah terbit	Penerbit Informatika Bandung
2	Publikasi Ilmiah Jurnal Internasional	accepted/published	Computer Speech and Language (Elsevier)

#### Luaran Tambahan

Tahun Luaran	Jenis Luaran	Status target capaian ( <i>accepted, published, terdaftar atau granted, atau status lainnya</i> )	Keterangan ( <i>url dan nama jurnal, penerbit, url paten, keterangan sejenis lainnya</i> )
2	Hak Cipta	terdaftar	Korpus video-teks Bahasa Indonesia dan model deep learning hasil pembelajaran

## 5. ANGGARAN

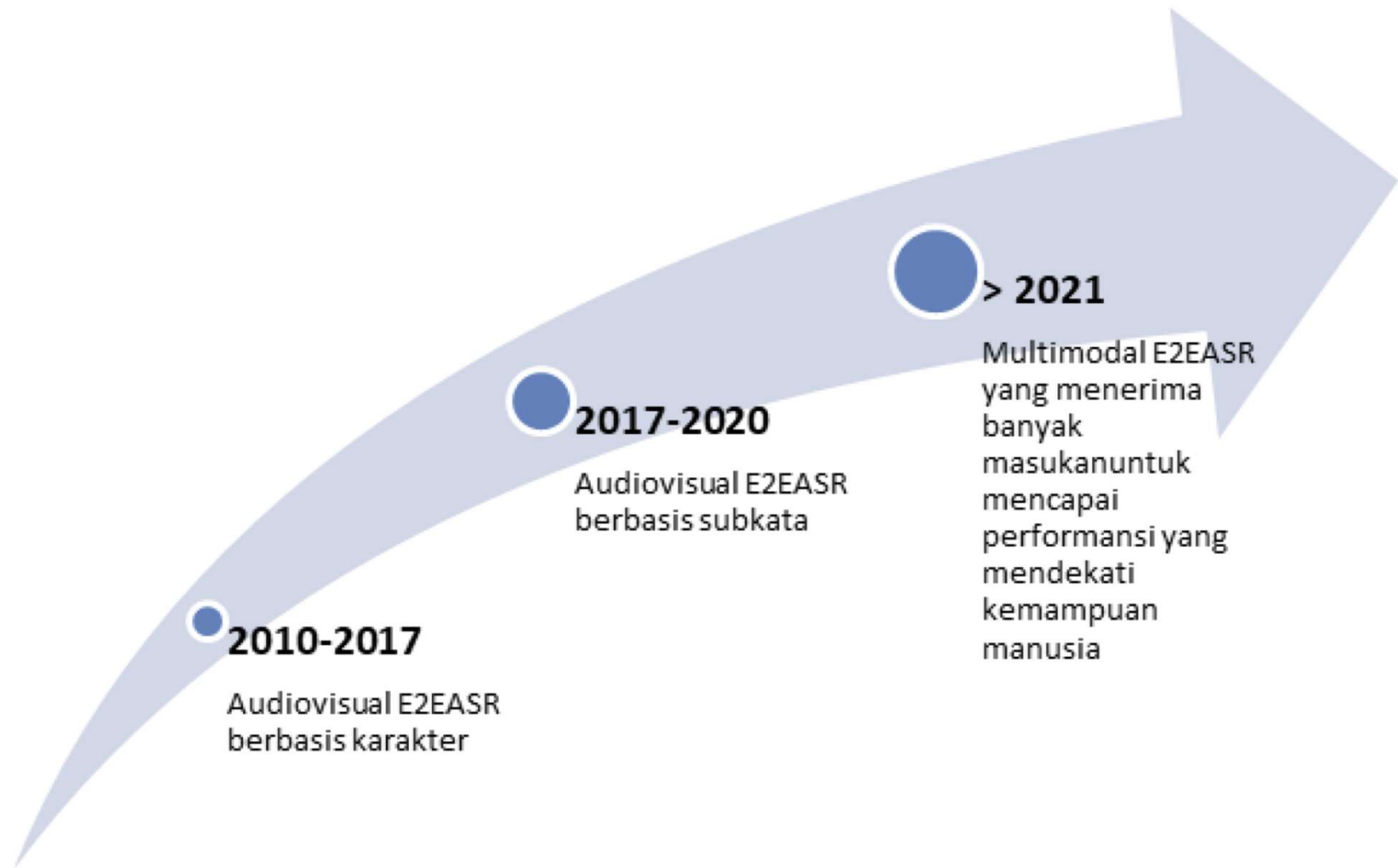
Rencana anggaran biaya PPM mengacu pada PMK yang berlaku dengan besaran minimum dan maksimum sebagaimana diatur pada buku Panduan Penelitian dan Pengabdian kepada Masyarakat Edisi 12.

**Total RAB 2 Tahun Rp. 92,500,000**

**Tahun 1 Total Rp. 41,000,000**

Jenis Pembelanjaan	Item	Satuan	Vol.	Biaya Satuan	Total
Bahan	ATK	Paket	1	1,500,000	1,500,000
Bahan	Barang Persediaan	Unit	1	4,000,000	4,000,000
Sewa Peralatan	Peralatan penelitian	Unit	1	2,000,000	2,000,000
Analisis Data	HR Sekretariat/Administrasi Peneliti	OB	1	300,000	300,000
Analisis Data	HR Pengolah Data	OP(penelitian)	1	1,540,000	1,540,000

Menurut prediksi Cisco, **IP video traffic pada tahun 2020 nanti diperkirakan akan mencapai 82%** dari semua *IP traffic* di seluruh dunia. Oleh sebab itu, *data mining* yang selama ini dilakukan terhadap basisdata dan teks mulai bergeser ke *video mining*. Teknik *video mining* tentu saja memerlukan *automatic speech recognition* (ASR) berbasis audiovisual untuk mengkonversi video ke teks. Para ahli telah mengusulkan banyak metode konversi video ke teks, yang dapat dikelompokkan ke dalam tiga kategori, yaitu: 1) pengenalan ucapan secara audio, 2) pengenalan ucapan secara visual, dan 3) pengenalan ucapan secara audiovisual. Cara ketiga, suara dan gerak bibir pemberbicara digunakan sebagai masukan bagi sistem pengenalan ucapan secara audiovisual. Sejumlah ahli telah membuktikan bahwa sistem pengenalan ucapan audiovisual memberikan akurasi lebih tinggi dibanding dua pendekatan lainnya. **Namun, masih terdapat tiga masalah yang perlu diselesaikan**, yaitu: 1) bagaimana mengkombinasikan fitur audio dan visual; 2) model akustik yang berbasis fonem maupun grafem/karakter/huruf ternyata kurang tahan terhadap derau; dan 3) ekstraksi fitur yang umumnya berbasis *Mel Frequency Cepstral Coefficient* (MFCC) dan *Gaussian Mixture Model* (GMM) memiliki kompleksitas tinggi. Banyak pakar telah mengusulkan beragam model pengenalan ucapan audiovisual berbasis *deep learning* untuk mengatasi ketiga masalah tersebut. Sayangnya, untuk bahasa Inggris model-model yang telah diusulkan masih memiliki banyak kelemahan. Proposal penelitian ini **ditujukan untuk membangun sebuah model Pengenalan Ucapan Audio-Visual berbasis Silabel Bahasa Indonesia (PUAVSI)** menggunakan pendekatan *Latent Sequence Decompositions* (LSD), yang berbasis subkata (deretan karakter/huruf, bukan satu karakter tunggal), dan menerima masukan sinyal ucapan mentah (tanpa perlu diekstraksi menggunakan MFCC). **Metode penelitian untuk membangun model PUAVSI terdiri atas tiga tahapan**, yaitu: 1) pembangunan korpus video; 2) augmentasi data suara dan video dan pembelajaran model LSD berbasis *deep learning* menggunakan gabungan fitur audio dan visual; 3) evaluasi model PUAVSI. Sebagian proses pada tahap pertama sudah selesai dikerjakan di tahun 2017 dengan hasil berupa korpus teks berisi tujuh ribuan kalimat (diekstrak dari 10 juta kalimat dikumpulkan dari berbagai situs berita) yang kaya *silabel* yang akan digunakan sebagai transkripsi yang dibaca para pembicara saat perekaman korpus video. Luaran utama penelitian ini adalah satu publikasi ilmiah di jurnal internasional ***Computer Speech and Language* (Elsevier)**, satu **buku ber-ISBN** yang diterbitkan oleh Penerbit Informatika Bandung sedangkan luaran tambahannya berupa **dua Hak Cipta** (yaitu: korpus video-teks Bahasa Indonesia dan model *deep learning* hasil pembelajaran). Pada tahun pertama ditargetkan menghasilkan sebuah model PUAVSI dengan TKT 2 dan pada tahun kedua diharapkan menghasilkan model PUAVSI yang telah diuji laboratorium dengan **TKT 3**.



**2010-2017**

Audiovisual E2EASR  
berbasis karakter

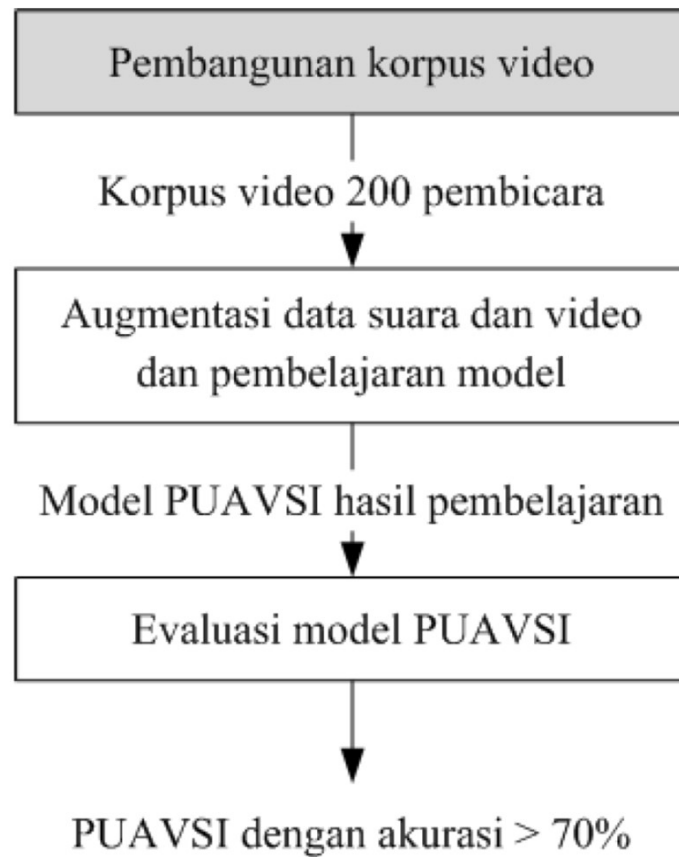
**2017-2020**

Audiovisual E2EASR  
berbasis subkata

**> 2021**

Multimodal E2EASR  
yang menerima  
banyak  
masukan untuk  
mencapai  
performansi yang  
mendekati  
kemampuan  
manusia







1. Cisco, "The Zettabyte Era - Trends and Analysis," **2016**. [Online]. Available: <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-indexvni-hyperconnectivity-wp.html>.
2. S. Dupont and J. Luetin, "Audio-visual speech modeling for continuous speech recognition," *IEEE Trans. Multimed.*, vol. 2, no. 3, pp. 141–151, Sep. 2000.
3. K. Noda, Y. Yamaguchi, K. Nakadai, H. G. Okuno, and T. Ogata, "Audio-visual speech recognition using deep learning," *Appl. Intell.*, vol. 42, no. 4, pp. 722–737, **2015**.
4. R. Janakiraman, J. C. Kumar, and H. A. Murthy, "Robust syllable segmentation and its application to syllable-centric continuous speech recognition," in *National Conference on Communications (NCC)*, 2010, pp. 1–5.
5. A. Andreo, "Audio-Visual Speech Recognition," Center for Language and Speech Processing, University, Johns Hopkins, **2017**. [Online]. Available: <https://www.clsp.jhu.edu/workshops/00-workshop/audio-visual-speech-recognition/>.
6. K. Saenko, K. Livescu, M. Siracusa, K. Wilson, J. Glass, and T. Darrell, "Visual speech recognition with loosely synchronized feature streams," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, 2005, vol. 2, p. 1424–1431 Vol. 2.
7. A. W. Liew, S. Wang, and K. Klinger, *Visual Speech Recognition: Lip Segmentation and Mapping*. IGI Global, 2009.
8. I. A. Eldirawy, "Visual Speech Recognition," Islamic University of Gaza, 2011.
9. A. Thanda and S. M. Venkatesan, "Audio Visual Speech Recognition using Deep Recurrent Neural Networks," *CoRR*, vol. abs/1611.0, **2016**.
10. H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, vol. 264, pp. 746–748, 1976.
11. D. Amodei, R. Anubhai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, J. Chen, M. Chrzanowski, A. Coates, G. Diamos, E. Elsen, J. Engel, L. Fan, C. Fougner, T. Han, A. Hannun, B. Jun, P. LeGresley, L. Lin, S. Narang, A. Ng, S. Ozair, R. Prenger, J. Raiman, S. Satheesh, D. Seetapun, S. Sengupta, Y. Wang, Z. Wang, C. Wang, B. Xiao, D. Yogatama, J. Zhan, and Z. Zhu, "Deep Speech 2: End-to-End Speech Recognition in English and Mandarin," *Cornell University Library's arXiv.org*, **2015**. [Online]. Available: <http://arxiv.org/abs/1512.02595>.
12. K. Arulkumar, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep Reinforcement Learning: A Brief Survey," *IEEE Signal Process. Mag. Spec. Issue Deep Learn. Image Underst.*, vol. 34, no. 6, pp. 26–38, **2017**.
13. C.-C. Chiu, T. N. Sainath, Y. Wu, R. Prabhavalkar, P. Nguyen, Z. Chen, A. Kannan, R. J. Weiss, K. Rao, E. Gonina, N. Jaitly, B. Li, J. Chorowski, and M. Bacchiani, "State-of-the-art Speech Recognition With Sequence-to-Sequence Models," *Cornell University Library's arXiv.org*, **2017**. [Online]. Available: <http://arxiv.org/abs/1712.01769>.
14. J. Chorowski and N. Jaitly, "Towards better decoding and language model integration in sequence to sequence models," *Cornell University Library's arXiv.org*, **2016**.
15. W. Chan, N. Jaitly, Q. Le, and O. Vinyals, "Listen, attend and spell: A neural network for large vocabulary conversational speech recognition," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, **2016**, pp. 4960–4964.
16. W. Chan and Y. Zhang, "Latent Sequence Decompositions," *Cornell University Library's arXiv.org*, **2017**.
17. I. Rebai, Y. Benayed, W. Mahdi, and J. P. Lorré, "Improving speech recognition using data augmentation and acoustic model fusion," *Procedia Comput. Sci.*, vol. 112, pp. 316–322, **2017**.
18. N. Zeghidour, N. Usunier, G. Synnaeve, R. Collobert, and E. Dupoux, "End-to-End Speech Recognition From the Raw Waveform," *Cornell University Library's arXiv.org*, **2018**. [Online]. Available: <http://arxiv.org/abs/1806.07098>.

## 1. JUDUL PENELITIAN

Silabifikasi dan Fonemisasi untuk Model Text-To-Speech Bahasa Indonesia yang Lebih Akurat dan Natural

Bidang Fokus RIRN / Bidang Unggulan Perguruan Tinggi	Tema	Topik (jika ada)	Rumpun Bidang Ilmu
Teknologi Informasi dan Komunikasi	-	Framework/platform penunjang industri kreatif dan kontrol	Ilmu Komputer

Kategori (Kompetitif Nasional/ Desentralisasi/ Penugasan)	Skema Penelitian	Strata (Dasar/ Terapan/ Pengembangan)	SBK (Dasar, Terapan, Pengembangan)	Target Akhir TKT	Lama Penelitian (Tahun)
Penelitian Desentralisasi	Penelitian Terapan Unggulan Perguruan Tinggi	SBK Riset Terapan	SBK Riset Terapan	6	2

## 2. IDENTITAS PENGUSUL

Nama, Peran	Perguruan Tinggi/ Institusi	Program Studi/ Bagian	Bidang Tugas	ID Sinta	H-Index
SUYANTO Ketua Pengusul	Universitas Telkom	Informatika		28705	13
Dr. ADE ROMADHONY S.T, M.T Anggota Pengusul 1	Universitas Telkom	Informatika	Pembangunan korpus suara	6018677	2
FEBRYANTI STHEVANIE S.T, M.T Anggota Pengusul 2	Universitas Telkom	Informatika	Evaluasi model	6026663	3

### 3. MITRA KERJASAMA PENELITIAN (JIKA ADA)

Pelaksanaan penelitian dapat melibatkan mitra kerjasama, yaitu mitra kerjasama dalam melaksanakan penelitian, mitra sebagai calon pengguna hasil penelitian, atau mitra investor

Mitra	Nama Mitra
Mitra Calon Pengguna	Teguh Eko Budiarto

#### 4. LUARAN DAN TARGET CAPAIAN

##### Luaran Wajib

Tahun Luaran	Jenis Luaran	Status target capaian ( <i>accepted, published, terdaftar atau granted, atau status lainnya</i> )	Keterangan ( <i>url dan nama jurnal, penerbit, url paten, keterangan sejenis lainnya</i> )
1	Dokumen pendaftaran hak cipta	Terbit Sertifikat	
2	Dokumen hasil uji	Ada/tersedia	Program komputer Indonesian Text-to-Speech yang diberi nama "INATTS"

##### Luaran Tambahan

Tahun Luaran	Jenis Luaran	Status target capaian ( <i>accepted, published, terdaftar atau granted, atau status lainnya</i> )	Keterangan ( <i>url dan nama jurnal, penerbit, url paten, keterangan sejenis lainnya</i> )
1	Artikel pada Conference/Seminar Internasional di Pengindeks Bereputasi	Terbit dalam Prosiding	ICoICT
2	Artikel di Jurnal Internasional Terindeks di Pengindeks Bereputasi	Accepted	International Journal of Speech Technology (Springer)

**Total RAB 2 Tahun Rp. 435,680,000**

**Tahun 1 Total Rp. 217,340,000**

<b>Jenis Pembelanjaan</b>	<b>Komponen</b>	<b>Item</b>	<b>Satuan</b>	<b>Vol.</b>	<b>Biaya Satuan</b>	<b>Total</b>
Bahan	ATK	ATK	Paket	1	4,000,000	4,000,000
Bahan	Bahan Penelitian (Habis Pakai)	Tinta printer	Paket	1	3,000,000	3,000,000
Bahan	Bahan Penelitian (Habis Pakai)	Perekaman kumpulan kalimat di dalam teks transkripsi yang berisi lima ribuan silabel	Paket	1	50,000,000	50,000,000
Bahan	Barang Persediaan	Hardisk SSD	Unit	1	7,000,000	7,000,000
Pengumpulan	FGD persiapan	FGD persiapan	Paket	1	10,000,000	10,000,000



Teknologi *text-to-speech* (TTS) atau sintesis ucapan untuk Bahasa Inggris maupun bahasabahasa lain yang memiliki sumberdaya tinggi (*high-resource languages*) umumnya sudah lumayan akurat dan terdengar **natural untuk kata-kata formal**. Namun, TTS seringkali menghasilkan ucapan yang **tidak tepat dan kurang natural untuk entitas nama**. Permasalahan ini menjadi lebih banyak ditemukan untuk bahasa-bahasa yang memiliki sumberdaya rendah (*low-resource languages*), seperti Bahasa Indonesia. Hal ini disebabkan oleh primitif dasar (basis) yang umum digunakan adalah *phoneme* atau *diphone*. Untuk mengatasi masalah tersebut, sejumlah ahli telah mengembangkan teknologi TTS berbasis silabel, yang mampu memberikan naturalitas yang lebih tinggi. **Pada penelitian ini, dilakukan analisis, desain, dan implementasi model TTS berbasis silabel untuk bahasa Indonesia yang diberi nama INATTS**. Dua model yang telah dihasilkan pada penelitian sebelumnya, yaitu Silabifikasi dan Fonemisasi, akan dimanfaatkan secara maksimal untuk menghasilkan model INATTS yang akurat dan natural untuk mengucapkan kata-kata formal, informal, maupun entitas nama. **Metode penelitian untuk membangun model INATTS terdiri atas tiga tahapan**, yaitu: 1) pembangunan korpus suara; 2) desain dan pembelajaran model INATTS; 3) evaluasi model INATTS. **Luaran utama penelitian** ini adalah hak cipta berupa program komputer “INATTS” dan dokumen hasil uji coba produk. Sementara itu, terdapat **dua luaran tambahan** yang berupa satu makalah ilmiah yang dipublikasikan di konferensi internasional ICoICT dan satu makalah ilmiah yang dipublikasikan di jurnal internasional bereputasi “International Journal of Speech Technology”, Springer, Scopus, Q2, SJR = 0.24, H-Index = 26.

1. S. Suyanto, et al., "Indonesian syllabification using a pseudo nearest neighbour rule and phonotactic knowledge," *Speech Communication*, Elsevier, vol. 85, pp. 109–118, 2016. <http://dx.doi.org/10.1016/j.specom.2016.10.009>
2. S. Suyanto, "Incorporating syllabification points into a model of grapheme-to-phoneme conversion," *Int. J. Speech Technology*, vol. 22, no. 2, pp. 459–470, Jun. 2019. <https://doi.org/10.1007/s10772-019-09619-4>
3. E. A. Parande and S. Suyanto, "Indonesian graphemic syllabification using a nearest neighbour classifier and recovery procedure," *Int. J. Speech Technology* (22) 1, pp. 13–20, 2019. <https://link.springer.com/article/10.1007/s10772-018-09569-3>
4. S. Suyanto, "Flipping onsets to enhance syllabification," *International Journal of Speech Technology*, Springer, vol. 22, no. 4, pp. 1031–1038, 2019. <https://link.springer.com/article/10.1007/s10772-019-09649-y>
5. S. Suyanto, "Phonological similarity-based backoff smoothing to boost a bigram syllable boundary detection," *International Journal of Speech Technology*, Springer, vol. 23, no. 1, pp. 191–204, 2020. <https://doi.org/10.1007/s10772-020-09677-z>
6. Suyanto, Kemas Muslim Lhaksana, Moch. Arif Bijaksana, and Adriana Kurniawan, "Data Augmentation Methods for Low-Resource Orthographic Syllabification", *IEEE Access*, Vol. 8, pp. 147399-147406, August 2020, Electronic ISSN: 2169-3536, DOI: <https://doi.org/10.1109/ACCESS.2020.3015778>
7. S. Suyanto, A. Sunyoto, R. N. Ismail, E. Rachmawati, and W. Maharani, "Stemmer and phonotactic rules to improve n-gram tagger-based Indonesian phonemicization," *Journal of King Saud University - Computer and Information Sciences*, no. xxxx, 2021, ISSN: 1319-1578, doi: <https://doi.org/10.1016/j.jksuci.2021.01.006>

#### Tahun 2021

1. SUO, Rezza, "Augmented-Syllabification ..." (26-05-2020, IJST Springer Q2)
2. SUO, ADE, FSV, Rezza, Bayu, "Dokumen pendaftaran hak cipta INATTS" (31-08-2021)

#### Tahun 2022

1. SUO, ADE, FSV, Rezza, "Indonesian TTS using FastSpeech ..." (Submitted 31-10-2021, Elsevier Q1)
2. ADE, FSV, SUO, Bayu, "To be proposed..."
3. FSV, ADE, SUO, Bayu, "To be proposed..."

**2016-2020**

Riset Dasar berupa desain, implementasi, model Silabifikasi, Fonemisasi, dan INASR dengan **TKT 3**

**2021-2023**

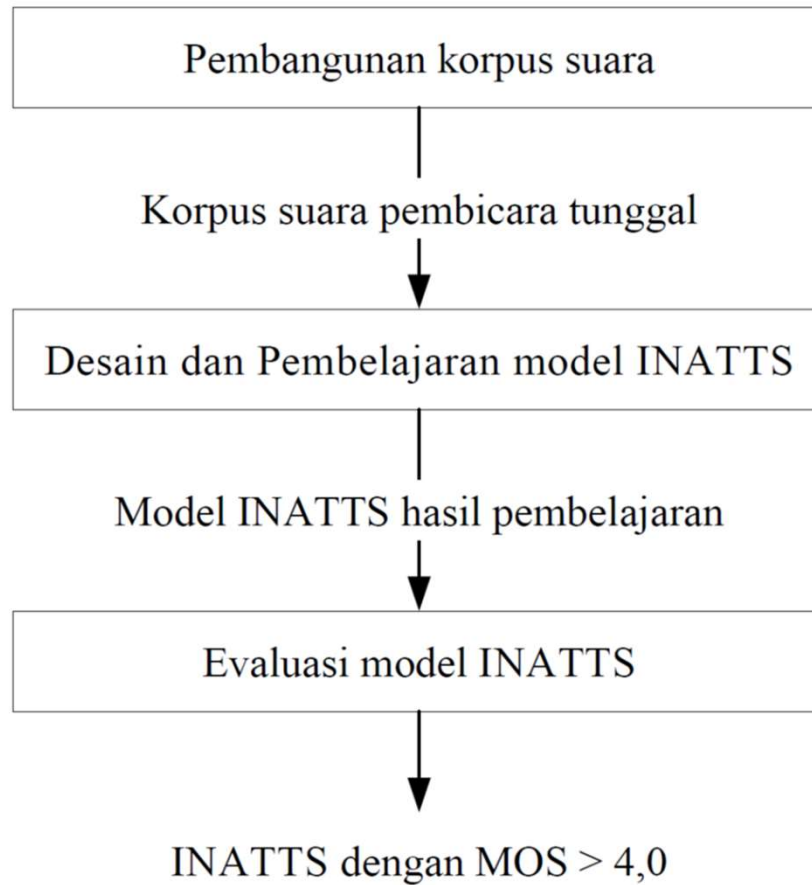
Riset Terapan pemanfaatan Silabifikasi, Fonemisasi, dan INASR untuk membangun model INATTS yang mencapai **TKT 6**

**2024-2026**

Riset Pengembangan untuk menghasilkan prototipe INATTS dengan **TKT 8**

**2027-2030**

Riset Pengembangan INATTS untuk aplikasi dunia nyata berskala besar di industri dengan **TKT 9**



**Gambar 2.** Metode penelitian yang diusulkan

## JADWAL

### Tahun ke-1

No	Nama Kegiatan	Bulan											
		1	2	3	4	5	6	7	8	9	10	11	12
1	Pembangunan korpus suara satu orang pria	■	■	■	■	■	■						
2	Desain model INATTS dengan memanfaatkan model INASR			■	■	■	■	■	■				
3	Implementasi dan proses pembelajaran model INATTS dengan suara pria					■	■	■	■	■	■		
4	Pengujian model INATTS dengan suara pria hingga mencapai TKT 4							■	■	■	■	■	■

### Tahun ke-2

No	Nama Kegiatan	Bulan											
		1	2	3	4	5	6	7	8	9	10	11	12
1	Pembangunan korpus suara satu orang wanita	■	■	■	■	■	■						
2	Optimasi desain model INATTS			■	■	■	■	■	■				
3	Implementasi dan proses pembelajaran model INATTS dengan suara pria dan wanita					■	■	■	■	■	■		
4	Pengujian model INATTS dengan suara pria dan wanita hingga mencapai TKT 6							■	■	■	■	■	■

- [1] H. working Group, "HMM/DNN-based speech synthesis system (HTS)," **2016**. <http://hts.sp.nitech.ac.jp/>.
- [2] Z. Wu and O. Watts, "Merlin: An Open Source Neural Network Speech Synthesis System," in *9th ISCA Speech Synthesis Workshop*, 2016, pp. 202–207.
- [3] T. Hayashi *et al.*, "Espnet-TTS: Unified, Reproducible, and Integratable Open Source Endto- End Text-to-Speech Toolkit," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2020, vol. 2020-May, pp. 7654–7658.
- [4] S. Watanabe *et al.*, "ESPnet : End-to-End Speech Processing Toolkit," *arXiv*, pp. 1–5, **2018**.
- [5] S. Karita *et al.*, "A comparative study on Transformer vs RNN in speech applications," *arXiv*, vol. 9, no. 4, **2019**.
- [6] S. **Suyanto**, A. Arifianto, A. Sirwan, and A. P. Rizaendra, "End-to-End Speech Recognition Models for a Low-Resourced Indonesian Language," in **2020 8th International Conference on Information and Communication Technology (ICoICT)**, Jun. 2020, pp. 1–6, doi: 10.1109/ICoICT49345.2020.9166346.
- [7] S. **Suyanto**, S. Hartati, A. Harjoko, and D. Van Compernelle, "Indonesian syllabification using a pseudo nearest neighbour rule and phonotactic knowledge," *Speech Commun.*, vol. 85, pp. 109–118, **2016**, doi: <http://dx.doi.org/10.1016/j.specom.2016.10.009>.
- [8] E. A. Parande and S. **Suyanto**, "Indonesian graphemic syllabification using a nearest neighbour classifier and recovery procedure," *Int. J. Speech Technol.*, vol. 22, no. 1, pp. 13– 20, **2019**, doi: <https://doi.org/10.1007/s10772-018-09569-3>.
- [9] S. **Suyanto**, "Flipping onsets to enhance syllabification," *Int. J. Speech Technol.*, vol. 22, no. 4, pp. 1031–1038, **2019**, doi: <https://doi.org/10.1007/s10772-019-09649-y>.
- [10] S. **Suyanto**, K. M. Lhaksana, M. A. Bijaksana, and A. Kurniawan, "Data Augmentation Methods for Low-Resource Orthographic Syllabification," *IEEE Access*, vol. 8, pp. 147399–147406, **2020**, doi: <https://doi.org/10.1109/ACCESS.2020.3015778>.
- [11] S. **Suyanto**, "Phonological similarity-based backoff smoothing to boost a bigram syllable boundary detection," *Int. J. Speech Technol.*, vol. 23, no. 1, pp. 191–204, **2020**, doi: <https://doi.org/10.1007/s10772-020-09677-z>.
- [12] R. N. Ismail and S. **Suyanto**, "Indonesian Graphemic Syllabification Using n-Gram Tagger with State-Elimination," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*, Jun. **2020**, pp. 1–6, doi: 10.1109/ICoICT49345.2020.9166368.
- [13] A. Kurniawan and S. **Suyanto**, "Syllable-Based Indonesian Lip Reading Model," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*, Jun. **2020**, pp. 1–6, doi: 10.1109/ICoICT49345.2020.9166217.

## Kekinian dan sumber primer pengacuan:

- < 30% (1 poin)
- 30 – 50% (2 poin)
- 50 – 80% (3 poin)
- > 80% (5 poin)

- [14] **Suyanto** and A. Harjoko, "Nearest neighbour-based Indonesian G2P conversion," *Telkomnika (Telecommunication, Comput. Electron. Control.*, vol. 12, no. 2, pp. 389–396, 2014, doi: <http://dx.doi.org/10.12928/telkomnika.v12i2.57>.
- [15] S. **Suyanto**, S. Hertanto, and A. Harjoko, "Modified Grapheme-Fragment and Phoneme Rule-Based Approach for Indonesian G2P Conversion," *Journal of Indonesian System*, no. 1, pp. 1–10, 2019.
- [16] S. **Suyanto**, S. Hertanto, and A. Harjoko, "PNNR-Based Indonesian G2P Conversion," *Journal of Indonesian System*, no. 1, pp. 459–470, Jun. 2019.
- [17] T. Okamoto, "Text-to-speech systems," in *Handbook of Text-to-Speech*, pp. 1–10, 2019.
- [18] L. Juvelin, "Adversarial Text-to-Speech," *arXiv preprint arXiv:1905.09861*, 2019.
- [19] J. Shen, "Speech synthesis," *Speech synthesis*, pp. 1–10, 2019.
- [20] A. Tama, "Text-to-speech synthesis," *Text-to-speech synthesis*, pp. 1–10, 2019.
- [21] N. Kalchbrenner, "Text-to-speech synthesis," *Text-to-speech synthesis*, pp. 1–10, 2019.
- [22] Y. Cao, "Text-to-speech synthesis," *Text-to-speech synthesis*, pp. 1–10, 2019.
- [23] Y. Wang, "Text-to-speech synthesis," *Text-to-speech synthesis*, pp. 1–10, 2019.
- [24] N. Li, S. Ren, and X. Han, "FastSpeech: Fast, robust and controllable text-to-speech," *arXiv preprint arXiv:1905.10255*, 2019.
- [25] Y. Ren and X. Han, "FastSpeech: Fast, robust and controllable text-to-speech," *arXiv preprint arXiv:1905.10255*, 2019.
- [26] C. O. S. E. L. Earning, W. Ping, K. Peng, A. Gibiansky, O. Sercan, and J. Raiman, "Deep Voice 3: Scaling text-to-speech with convolutional sequence learning," in *ICLR*, 2018, pp. 1–16.

*Peripapillary atrophy* (PPA) adalah salah satu indikasi glaukoma yang dapat diamati pada citra fundus. **Glaukoma adalah penyakit mata penyebab kebutaan terbesar kedua di dunia dan merupakan penyakit mata yang tidak dapat disembuhkan.** Oleh karena itu, PPA perlu dideteksi kemunculannya sehingga penyakit tersebut dapat diketahui lebih awal. Selain itu, PPA juga merupakan salah satu indikasi dari myopia atau rabun jauh yang umumnya diderita tidak hanya oleh orang dewasa tetapi juga anak-anak. PPA merupakan bagian dari retina, di mana kemunculannya berada di luar area yang berbentuk lingkaran yang disebut dengan *optic nerve head* dengan tekstur yang khas. Kemunculan PPA pada citra fundus dapat memiliki tekstur yang tampak samar dan halus (*mild*) atau jelas dan tegas (*severe*). *Ophthalmologist* sebenarnya dapat mendeteksi kemunculan dari PPA, tetapi PPA dengan struktur *mild* dan banyaknya data akan membutuhkan waktu deteksi yang lama serta perbedaan latar belakang dari *ophthalmologist* dapat mengakibatkan hasil deteksi menjadi subjektif. **Pada penelitian sebelumnya pendeteksian PPA hanya dikelompokkan menjadi dua kelas yaitu ada atau tidak ada PPA.** Hasil tersebut perlu dikembangkan karena struktur PPA *mild* dan *severe* dapat menunjukkan perbedaan tingkat keparahan glaucoma. **Pada penelitian ini diusulkan metode klasifikasi PPA baru yang mampu mengelompokkan PPA menjadi tiga kelas yaitu tidak ada (*no-PPA*), *mild-PPA*, dan *severe-PPA*.** Selain itu, metode segmentasi PPA juga akan diusulkan untuk menunjukkan area dari kemunculan PPA. **Metode klasifikasi PPA yang diusulkan terdiri atas dua tahap** yaitu *training* dan *testing*. Pada *training* menerima masukan citra fundus *training* beserta label kelas terdiri dari *no/mild/severe*, sedangkan pada *testing* menerima citra fundus *testing* dan data klinis sebagai *ground truth* untuk pengujian. Kedua tahap tersebut melakukan tiga proses yang sama, yaitu: 1) pembentukan ROI untuk membentuk sub-citra dengan fokus pada area PPA; 2) *pre-processing* untuk meningkatkan kualitas citra dengan melakukan penyesuaian ruang warna dan mencoba beberapa metode *image enhancement* untuk mengetahui metode yang paling sesuai; dan 3) ekstraksi ciri untuk menghasilkan nilai ciri menggunakan suatu metode berdasarkan proses uji coba beberapa metode ekstraksi ciri untuk menentukan metode yang paling tepat. Selanjutnya, pada tahap *training* berbasis *Dynamic Ensemble Selection* (DES) akan dilakukan seleksi ciri untuk menentukan ciri hasil ekstraksi yang memiliki peran penting pada proses klasifikasi yang dilakukan pada tahap *testing*. **Luaran wajib** penelitian ini adalah satu makalah ilmiah per tahun yang dipublikasikan di jurnal internasional bereputasi terindeks Scopus “*Computerized Medical Imaging and Graphics*”, Elsevier, Q1, SJR = 1.04, H-Index = 70. Sementara itu, terdapat **luaran tambahan** yang berupa hak cipta basis data citra fundus berlabel di tahun pertama dan hak cipta program komputer model deteksi glaukoma multikelas dengan TKT 3.

Pembacaan gerak bibir atau *lip reading* adalah sebuah metode komunikasi untuk memahami materi yang diucapkan seseorang, dengan membaca gerakan bibir dari pembicara. Dalam banyak literatur, pembacaan gerak bibir seringkali disebut *Visual Speech Recognition* atau pengenalan ucapan secara visual. Pengenalan ucapan secara visual hanya memanfaatkan fitur visual (video) untuk menghasilkan teks dengan mengabaikan fitur audio (suara) didalamnya. Teks yang dihasilkan terdiri dari beberapa kata atau bahkan kalimat yang diucapkan oleh pembicara. **Metode ini tidak hanya dapat digunakan untuk membantu orang yang kehilangan kemampuan pendengaran, namun juga orang yang kehilangan kemampuan berbicara. Tidak hanya terbatas disana, metode ini juga dapat membantu teknologi pengenalan suara. Dalam kondisi lingkungan dengan tingkat kebisingan yang tinggi, akan sulit untuk dilakukan pengenalan suara.** Sehingga metode pembacaan gerak bibir akan sangat membantu dalam pengenalan pengucapan. **Namun masih terdapat beberapa tantangan yang harus diselesaikan** dalam pembacaan gerak bibir seperti : 1) Bagaimana mengatasi tingginya variasi input fitur visual (bentuk wajah, bentuk bibir, bekas luka, dan fitur lain yang terdapat di wajah), 2) Bagaimana mengatasi perbedaan kecepatan pembicara yang mengakibatkan input visual menjadi bervariasi meskipun mengucapkan kata yang sama, 3) Adanya kemungkinan besar *Out of Vocabulary* (OOV), atau ketidakmampuan pengenalan kata disebabkan kata tersebut diluar dari yang telah dikenali (tidak pernah dipelajari sebelumnya). Pada sebagian besar kasus yang terjadi, peneliti akan membatasi inputan dan mengontrol penuh lingkungan pengambilan data. Namun hal tersebut akan membuatnya menjadi berjalan secara tidak natural. Sistem yang tidak berjalan secara natural akan sulit diterapkan secara nyata. Seiring perkembangan kemampuan komputasi dan teknologi, terdapat sebuah metode yang cukup mumpuni untuk mengatasi perbedaan fitur-fitur yang telah disebutkan yaitu *deep learning*. *Deep Learning* memberikan hasil yang menjanjikan dalam mengekstraksi fitur visual. Ini akan memberikan kesempatan bagi *lip reading* untuk memberikan hasil yang lebih baik. Terhadap masalah OOV, penggunaan pengenalan berbasis silabel (suku kata) akan menangani masalah tersebut. Sistem pengenalan berbasis suku kata akan memberikan kemampuan untuk membangun kata baru yang tidak pernah dipelajari sebelumnya. Kombinasi suku kata yang pernah dipelajari sebelumnya akan digunakan untuk membangun kata baru. **Proposal penelitian ini ditujukan untuk membangun sebuah model pembacaan gerak bibir berbasis silabel Bahasa Indonesia menggunakan Deep Learning mampu menangani OOV. Luaran utama penelitian ini adalah satu publikasi ilmiah di konferensi internasional yang terindeks SCOPUS.**



Silabifikasi otomatis adalah proses membagi kata menjadi deretan silabel. **Model silabifikasi sangat penting untuk beberapa penelitian dan pengembangan aplikasi di bidang linguistik, seperti sintesis pidato, pengenalan ucapan, pengenalan emosi bicara, identifikasi dialek, mesin penerjemah, pemeriksa ejaan, temu-kembali informasi, dan banyak lagi aplikasi lainnya.** Silabifikasi kata dan entitas nama Bahasa Indonesia memiliki tingkat ambiguitas yang tinggi. Saat ini, model silabifikasi entitas nama dan model silabifikasi kata-kata formal dibangun umumnya secara terpisah karena perbedaan karakteristik masalah yang relatif besar. **Hal ini membuat kedua model tersebut membutuhkan model lain, yaitu rekognisi entitas nama (*named-entity recognition*) sehingga kurang praktis.** Oleh karena itu, pada proposal penelitian ini diusulkan pembangunan tiga model: silabifikasi grafemis, fonemisasi, dan silabifikasi fonemis Bahasa Indonesia yang mampu melakukan silabifikasi dan fonemisasi kata-kata formal Kamus Besar Bahasa Indonesia (KBBI) sekaligus entitas nama dengan akurasi tinggi. Penelitian ini memiliki tingkat urgensi yang tinggi mengingat ketiga model tersebut sangat diperlukan dalam sistem pengenalan ucapan, baik secara audio, visual, maupun audiovisual, untuk berbagai kepentingan di perusahaan dan lembaga negara di Indonesia dalam bidang ekonomi, bisnis, keuangan, pers, politik, hingga keamanan negara. Lebih jauh, dengan sistem pengenalan ucapan Bahasa Indonesia yang handal negara akan dapat menjaga kedaulatan data dan memanfaatkan data secara optimal. membuat **Penelitian ini dirancang selama tiga tahun, dimulai sejak tahun 2020 hingga 2022. Targetnya,** pada akhir tahun 2022 akan dihasilkan model silabifikasi fonemis Bahasa Indonesia yang berakurasi tinggi. Penelitian ini sangat mendukung Rencana Induk Penelitian (RIP) Universitas Telkom yang akan memasuki masa Internasionalisasi pada periode 2022–2025, di mana Universitas Telkom akan berperan aktif dalam berbagai aktivitas tingkat regional (Asia Tenggara). **Luaran utama penelitian ini adalah setiap tahun dihasilkan satu publikasi ilmiah di jurnal internasional bereputasi yang termasuk 200 terbaik (Q1), yaitu *Computational Linguistic (MIT Press)* dan *Speech Communication (Elsevier)*, dengan status *accepted*. Selain itu, direncanakan terdapat luaran tambahan yang berupa tiga hak cipta, yang berupa basis data silabifikasi grafemis, basis data fonemisasi, dan basis data silabifikasi fonemis dengan status *granted* (telah bersertifikasi), serta sebuah paten yang berupa Model Silabifikasi Fonemis Bahasa Indonesia dengan status *registered* (telah terdaftar).**

Dengan jumlah pelanggan Indihome lebih dari empat juta di **pertengahan tahun 2018 ini dan diproyeksikan mencapai lima juta** di akhir tahun, PT Telkom perlu memberikan **layanan pelanggan (customer service) yang prima, cepat, dan murah**. Dua teknologi penting saat ini yang bisa mewujudkan hal itu adalah pengenalan ucapan otomatis atau *Automatic Speech Recognition* (ASR) dan *Machine Learning* (ML). ASR berfungsi mentranskripsi suara komplain pelanggan secara otomatis menjadi teks dan ML mengklasifikasikan teks tersebut ke dalam sejumlah kategori komplain untuk segera dilakukan penanganan. Sayangnya, hampir semua sistem ASR yang ada saat ini mengharuskan pembicara dekat dengan mikrofon dan dalam lingkungan yang rendah derau agar memberikan performansi yang tinggi. Hal ini tentu saja kurang praktis untuk berbagai aplikasi transkripsi suara, seperti layanan komplain pelanggan Indihome. Oleh karena itu, di sini diusulkan **penelitian yang bertujuan membangun model ASR yang tahan derau untuk Aplikasi Transkripsi Suara dan Klasifikasi Komplain (ATSKK)** pada Layanan Pelanggan Indihome. Secara ringkas **metode penelitian untuk membangun ATSKK terdiri atas lima tahapan**, yaitu: 1) Analisis dan desain ATSKK, 2) Pembangunan korpus suara, 3) Augmentasi data dan pembelajaran model ATSKK berbasis *deep learning*, 4) Pengujian ATSKK, dan 5) Pembangunan prototipe ATSKK. Sebagaimana proses pada tahap kedua sudah selesai dikerjakan pada tahun 2017, yaitu berupa pembangunan korpus teks berisi empat ribuan kalimat (diekstrak dari 10 juta kalimat dikumpulkan dari situs berita) yang kaya *triphone* (deretan tiga fonem kontekstual) yang akan digunakan sebagai transkripsi yang dibaca para pembicara saat perekaman korpus suara. **Luaran utama** penelitian ini adalah sebuah **model ATSKK dan sebuah dokumen feasibility study** sedangkan **luaran tambahannya** berupa satu publikasi ilmiah di jurnal internasional *Computer Speech and Language (Elsevier)*, **satu buku ber-ISBN** yang diterbitkan oleh Penerbit Informatika, Bandung, dan **dua Hak Cipta** (Korpus suara-teks Bahasa Indonesia dan Model *Deep Learning* hasil pembelajaran). Pada tahun pertama ditargetkan menghasilkan sebuah model ATSKK dengan **TKT 4** dan pada tahun kedua diharapkan menghasilkan sebuah prototipe ATSKK dengan **TKT 6**.



**Fakultas Informatika**  
School of Computing  
Telkom University

**Terima Kasih**